

Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity

Marianna Pensky

Department of Mathematics, University of Central Florida

Abstract

The present paper considers the problem of estimating a linear functional $\Phi = \int_{-\infty}^{\infty} \varphi(x)f(x)dx$ of an unknown deconvolution density f on the basis of n i.i.d. observations, Y_1, \dots, Y_n of $Y = \theta + \xi$, where ξ has a known pdf g , and f is the pdf of θ . Although various aspects and particular cases of this problem have been treated by a number of authors, the minimax theory for estimation of Φ has not been developed so far. In particular, there are no minimax lower bounds for an estimator of Φ in the case of an arbitrary function φ . The general upper bounds for the risk cover only the case when the Fourier transform of φ exists, and cannot be automatically applied to the case when φ is not absolutely or square integrable. Moreover, no theory exists for estimating Φ in the case when the vector of observations $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ is sparse. In addition, until now, the related problem of estimation of functionals $\Phi_n = n^{-1} \sum_{i=1}^n \varphi(\theta_i)$ in indirect observations has been treated as a separate problem with no connection to estimation of Φ .

The objective of the present paper is to fill in these gaps, to develop the general minimax theory of estimation of Φ , and to relate this problem to estimation of Φ_n . We offer a general, Fourier transform based approach to estimation of Φ (and Φ_n) and provide upper and minimax lower bounds for the risk in the case when function φ is square integrable. Furthermore, using technique of inversion formulas, we extend the theory to a number of situations when the Fourier transform of φ does not exist, but Φ can be presented as a functional of the Fourier transform of f and its derivatives. The latter enables us to construct minimax estimators of the functionals that have never been handled before such as the odd absolute moments and the generalized moments of the deconvolution density. Finally, we generalize our results to the situation when the vector $\boldsymbol{\theta}$ is sparse and the objective is estimating Φ (or Φ_n) over the nonzero components only. As a direct application of the proposed theory, we automatically recover multiple recent results and obtain a variety of new ones such as estimation of the mixing cumulative distribution function, estimation of the mixing probability density function with classical and Berkson errors and estimation of the $(2M+1)$ -th absolute moment of the deconvolution density.

Keywords and phrases: linear functional, minimax lower bound, sparsity, deconvolution

AMS (2000) Subject Classification: Primary: 62G20. Secondary: 62G05, 62G07

1 Introduction

In the present paper we consider the problem of estimating a linear functional

$$\Phi = \int_{-\infty}^{\infty} \varphi(x)f(x)dx \quad (1.1)$$

of an unknown deconvolution density f on the basis of observations Y_i , $i = 1, \dots, n$, where

$$Y_i = \theta_i + \xi_i, \quad i = 1, \dots, n. \quad (1.2)$$

Here, θ_i are i.i.d. random variables with unknown pdf f , and ξ_i are i.i.d. random errors with a known pdf g . The variable Y in this case is the mixture of θ and ξ , and the density f is sometimes referred to as the *mixing* density. Vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ in (1.2) may be sparse in the sense that, on the average, it has only k_n non-zero elements where $k_n/n \rightarrow 0$ as $n \rightarrow \infty$.

Note that the problem of estimating Φ in (1.1) appears in many contexts. If $\varphi(x) = \delta(x - x_0)$, then Φ is the value of the unknown deconvolution density f at the point x_0 , estimation of which has been studied extensively by Butucea and Comte (2009). If $\varphi(x) = \mathbb{I}(x < x_0)$, where $\mathbb{I}(\Omega)$ denotes the indicator of a set Ω , then problem (1.1) reduces to estimation of the mixing distribution function $\Phi = F(x_0)$ at x_0 examined by Dattner *et al.* (2011). If $\varphi(x) = e^{i\omega_0 x}$, then $\Phi = \hat{f}(\omega_0)$, the characteristic function of the mixing distribution at $\omega = \omega_0$. If $\varphi(x) = f_\eta(x - x_0)$, where f_η is a known pdf, then $\Phi = \Phi(x_0)$ is itself a convolution density at a point x_0 , as considered by Delaigle (2007). Finally, if $\varphi(x) = x^k$ or $\varphi(x) = |x|^{2M+1}$, then Φ is, respectively, the k -th moment or the $(2M+1)$ -th absolute moment of the mixing density f .

In addition, the problem of estimating Φ in (1.1) can be related to estimating of functionals in indirect observations

$$\Phi_n = \frac{1}{n} \sum_{i=1}^n \varphi(\theta_i). \quad (1.3)$$

A particular case of this problem, with $\varphi(x) = |x|$ and ξ_i being i.i.d. standard Gaussian errors, has been recently considered by Cai and Low (2011). Problems of the form (1.3) appear whenever one is interested in estimation of some characteristic which cannot be observed directly. For example, it is well known in nutritional research (see, e.g., Blundell (1998)), that people routinely miss-report their caloric intake as well as quantities of various food items which they consume. In order to accurately estimate the percentage of calories coming from, for example, fruit and vegetables, or the probability that a person consumes 25% more calories than the recommended amount, one needs to take these random errors into account. Similar situations occur in astronomy, where all measurements are subject to high levels of instrumental noise (see, e.g. Jaffe (2010)).

Indeed, if θ_i , $i = 1, \dots, n$ are i.i.d. with the pdf f and $\mathbb{E}|\varphi(\theta)| < \infty$, then Φ_n in (1.3) can be viewed as an “estimator” of $\Phi = \mathbb{E}\varphi(\theta)$ in (1.1) on the basis of “observations” θ_i . Moreover, as long as $\mathbb{E}|\varphi(\theta)|^2 < \infty$, one has $\mathbb{E}(\Phi_n - \Phi)^2 \leq n^{-1}\mathbb{E}|\varphi(\theta)|^2$. Hence, the minimax risks for estimating Φ_n and Φ are equivalent up to the Cn^{-1} additive term. Therefore, the upper bounds and the minimax lower bounds for the risks of both estimators will coincide up to, at most, a constant factor.

In the case when vector $\boldsymbol{\theta}$ is sparse and has, on the average, only k_n non-zero components, observations (1.2) can be viewed as heterogeneous sparse mixture, which was studied by a number of authors (see, e.g., Donoho and Jin (2004), Cai *et al.* (2007), Hall and Jin (2010) among others). Models of this type appear in various areas of applications such as early detection of bio-weapons use, detection of covert communications, meta-analysis with heterogeneity (see, e.g., Donoho and Jin (2004)), or detecting stellar occultations by Kuiper Belt objects (see, e.g., Liang *et al.* (2004)). In particular, in the case when ξ_i are i.i.d. Gaussian errors, research has been focused on testing multiple hypotheses that $\theta_i = 0$, and estimating non-zero components θ_i , $i = 1, \dots, n$.

If the signal is present ($k_n > 0$) and k_n is known, the problem of interest is to estimate some characteristics of nonzero elements of $\boldsymbol{\theta}$. The latter problem can be summarized as the problem of estimating

$$\Phi_{k_n} = \frac{1}{k_n} \sum_{i=1}^n \varphi(\theta_i) \mathbb{I}(\theta_i \neq 0). \quad (1.4)$$

Cai and Low (2011) considered estimation of (1.4) when $\varphi(x) = |x|$, the errors are Gaussian and $k_n = n^\nu$, $0 < \nu < 1$. They concluded that consistent estimation is impossible if $\nu \leq 1/2$, while

estimation yields the same minimax rates as in the non-sparse case for $\nu > 1/2$. The question of interest is whether the same will happen in general, or whether this phenomenon is due to the type of the functional (the first absolute moment) or the type of errors (Gaussian) studied in the paper. Note that, in the case of sparse vector $\boldsymbol{\theta}$, the pdf f of θ_i can be written as

$$f(x) = \mu_n f_0(x) + (1 - \mu_n) \delta(x), \quad \mu_n = k_n/n = n^{\nu-1},$$

where $f_0(\theta)$ is pdf of the nonzero entries of $\boldsymbol{\theta}$, and Φ_{k_n} in (1.4) corresponds to

$$\Phi_\mu = \int_{-\infty}^{\infty} \varphi(x) f_0(x) dx. \quad (1.5)$$

In spite of its great importance, surprisingly, the general problem of estimation of a linear functional of the deconvolution density has not been thoroughly investigated. In the non-sparse case, the problem of estimation of linear functional (1.1) of the mixing density with a square integrable function φ has been addressed by Butucea and Comte (2009) who derived the upper bounds for the mean squared risk for a variety of estimation scenarios, and constructed adaptive estimators that attain them (up to, at most, a logarithmic factor). However, the minimax lower risk bounds have been derived only in the case when $\varphi(x) = \delta(x - x_0)$ due to technical difficulties. Hence, although it is intuitively clear that the estimators constructed by Butucea and Comte (2009) attain (up to, at most, a logarithmic factor) the minimax lower bounds for the risk for other choices of $\varphi(x)$, to the best of our knowledge, this has never been proved. Moreover, the general study of Butucea and Comte (2009) has been confined to the situation when Fourier transform of function $\varphi(\theta)$ exists, which does not allow one to apply the theory when Φ is, for example, a mixing cdf at a point x_0 . Furthermore, as far as we know, there has never been a study of estimation of a linear functional of the deconvolution density in the sparse setting. For example, one would like to know for which values of ν one can estimate functionals (1.4) and (1.5) consistently and what the optimal convergence rate would be. The existing results for estimating of linear functionals in a general inverse problems setting, derived by Goldenshluger and Pereverzev (2000) and Mathé and Pereverzev (2002) focus on the upper bounds for the risk, do not involve the sparse setting and, in addition, cannot be easily adapted to estimation of functionals (1.1) or (1.3).

The purpose of the present paper is to fill in the existing gaps and to advance the theory of estimation of linear functionals of the deconvolution density. In particular, the paper accomplishes several key goals:

1. Derivation of minimax lower bound for the risk of an estimator of a general linear functional of the deconvolution density. These bounds have not been obtained before and they confirm that the estimators obtained via Fourier transform are indeed minimax optimal.
2. Estimation of linear functionals (1.1) when function φ is not necessarily integrable or square integrable using inversion formulas. This is a completely new technique.
3. Application of those methodologies to estimation of functionals of the form (1.3), which allows one to obtain estimators of those functionals with no additional effort. This part, although technically very simple, is philosophically important.
4. Estimation of the linear functionals of the form (1.5) (or (1.4)) in the case when deconvolution density f (or vector $\boldsymbol{\theta}$) is sparse. The advantage of the approach in this paper is that, by introducing an effective sample size and bias correction, we largely reduce this problem to estimation of a linear functional in a general set up, thus, significantly advancing the existing theory.

Below we provide a more detailed overview of the paper. We start the paper with the study of the case when Fourier transform of function $\varphi(x)$ in (1.1) exists in a regular sense. We refer to this situation as the **standard case** in comparison with the situations considered in Section 3 where Φ is represented as a linear functional Fourier transform f^* of f using some inversion formula. We complete the theory of Butucea and Comte (2009) by providing the exact expressions for the upper bounds for the risk in the case when f belongs to a Sobolev ball (Section 2.2) and establish the matching minimax lower bounds for the risk for a general function $\varphi(x)$ (Section 2.4). This confirms that the estimators derived in Butucea and Comte (2009) are indeed asymptotically optimal (or near optimal up to a logarithmic factor). In Section 2.3, we also explain how one can carry out an adaptive choice of the bandwidth using Lepskii's method. As an application of our methodology, we consider pointwise estimation of the mixing density with classical and Berkson errors (Section 2.5), thus, advancing the theory developed by Delaigle (2007).

Next, we expand our approach to incorporate estimation of functionals of the form (1.1) where function $\varphi(x)$ does not have the Fourier transform in a regular sense. In this case, functional Φ can often be represented via the Fourier transform of the deconvolution density or its derivatives, using technique of inversion formulas (Section 3). In Section 3.1 we provide a general formulation and some examples of the functionals that can be reduced to this form. Section 3.2 deals with the construction of estimators of linear functionals of the real and the imaginary parts of the Fourier transforms of derivatives of the deconvolution density. It also evaluates the risk bounds for those estimators. Section 3.3 provides examples of construction of functionals by using this methodology. In particular, we construct the estimators for the odd absolute moments of a deconvolution density, and also of a functional of the form $\int_{-\infty}^{\infty} \theta^m (\theta^2 + 1)^{-1} f(\theta) d\theta$ with $m \geq 2$. Minimax upper and lower bounds for the risks of those estimators are also derived. Also, as a result of application of the general theory, we immediately obtain an estimator of the mixing cdf studied by Dattner *et al.* (2011) and the first absolute moment of the mixing density, and the as well as the minimax lower and upper bounds for the risks of those estimators. The latter example significantly advances the theory developed in Cai and Low (2011) by generalizing it to the case of non-Gaussian errors and the mixing densities of various degrees of smoothness.

Section 4 deals with the situation where vector θ or deconvolution density f is sparse. We propose a general procedure designed for estimating functionals (1.4) and (1.5) in the sparse case for any function $\varphi(x)$ and any kind of error density g , and evaluate the upper bounds for the risk over Sobolev classes (Section 4.1). We construct the matching (up to, at most, a logarithmic factor) minimax lower bounds for the risk over those classes (Section 4.2). We discover that convergence rates in this case are determined by the “effective” sample size $n^{-1}k_n^2 = n^{2\nu-1}$ if $k_n = n^\nu$. The latter proves that conclusion of Cai and Low (2011) that consistent estimation is impossible if $\nu \leq 1/2$ applies not only to their particular case ($\varphi(x) = |x|$, Gaussian errors) but to any functional and any distribution of errors.

Finally, in Section 5 we carry out a finite sample simulation study of estimation of the first absolute moment of the mixing density. In particular, we compare the Fourier transform based estimator developed in the paper to the estimator of Cai and Low (2011). We show that, although estimators have similar precisions, the advantage of the estimators developed in the present paper is that they are adaptive to the choice of parameters and the smoothness of the unknown deconvolution density. The paper ends with a discussion in Section 6.

Although, in this paper, we restrict our attention to the case when the Fourier transform $f^*(\omega)$ of the unknown mixing density f has polynomial decay as $|\omega| \rightarrow \infty$, with some additional effort, the results in the paper can be extended to the case when $f^*(\omega)$ has exponential decay. This, however, is a matter for future consideration.

2 The minimax upper and lower bounds for the risk: the standard case

2.1 Notations and assumptions

For any function $t(x)$, we denote its Fourier transform by

$$t^*(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} t(x) dx.$$

Denote the pdf of observation Y_i by $q(y)$, so that $q^*(\omega) = f^*(\omega)g^*(\omega)$.

We assume that the mixing density belongs to the Sobolev ball $f \in \Omega_s(B)$ where

$$\Omega_s(B) = \left\{ t^* : \int_{-\infty}^{\infty} |t^*(\omega)|^2 (\omega^2 + 1)^s d\omega \leq B^2 \right\}, \quad s \geq 0. \quad (2.1)$$

We introduce the following standard assumptions on the known functions g and φ .

A1. There exist non-negative constants C_{g1} , C_{g2} , α , β and γ such that

$$|g^*(\omega)| \geq C_{g1}(\omega^2 + 1)^{-\alpha/2} \exp(-\gamma|\omega|^\beta), \quad (2.2)$$

$$|g^*(\omega)| \leq C_{g2}(\omega^2 + 1)^{-\alpha/2} \exp(-\gamma|\omega|^\beta), \quad (2.3)$$

where $\alpha > 0$ and $\beta = 0$ whenever $\gamma = 0$.

A2. There exist non-negative constants $C_{\varphi1}$, $C_{\varphi2}$, a , b and d such that

$$|\varphi^*(\omega)| \geq C_{\varphi1}(\omega^2 + 1)^{-a/2} \exp(-d|\omega|^b), \quad (2.4)$$

$$|\varphi^*(\omega)| \leq C_{\varphi2}(\omega^2 + 1)^{-a/2} \exp(-d|\omega|^b), \quad (2.5)$$

where $b = 0$ whenever $d = 0$.

In what follows, we use the symbol C for a generic positive constant, which takes different values at different places and is independent of n . Also, for any positive functions $a(n)$ and $b(n)$, we write $a(n) \asymp b(n)$ if the ratio $a(n)/b(n)$ is bounded above and below by finite positive constants independent of n .

2.2 Estimation and the upper bounds for the risk

In Section 2, we assume that the functional Φ in (1.1) can be represented as

$$\Phi = \frac{1}{2\pi} \int_{-\infty}^{\infty} f^*(\omega) \varphi^*(-\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{q^*(\omega)}{g^*(\omega)} \varphi^*(-\omega) d\omega \quad (2.6)$$

where the integral is absolutely convergent. This happens if, for example, $a > 1$ in (2.4), so that $|\varphi^*|$ and $|\varphi|$ are square integrable, however, this is true for a wider variety of functions φ (e.g., $\varphi(x) = \delta(x - x_0)$ considered in Butucea and Comte (2009)). We refer to this situation as

the **standard case** in comparison with the situations considered in Section 3 where Φ cannot be represented in the form (2.6).

Following Butucea and Comte (2009), we estimate Φ in (2.6) by

$$\widehat{\Phi}_h = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\widehat{q}^*(\omega)}{g^*(\omega)} \varphi^*(-\omega) \mathbb{I}(|\omega| \leq h^{-1}) d\omega \quad (2.7)$$

where

$$\widehat{q}^*(\omega) = n^{-1} \sum_{j=1}^n e^{i\omega Y_j} \quad (2.8)$$

is the unbiased estimator of $q^*(\omega)$ and $h = 0$ if function $|\varphi^*(\omega)|/|g^*(\omega)|$ has finite L^2 -norm. In particular, the upper bound for the risk of the estimator $\widehat{\Phi}_h$ over the Sobolev class $\Omega_s(B)$

$$R_n(\widehat{\Phi}_h, \Omega_s(B)) = \sup_{f \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_h - \Phi)^2$$

is given by the following inequality

$$R_n(\widehat{\Phi}_h, \Omega_s(B)) \leq \frac{\|g\|_{\infty}}{2\pi n} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2} \mathbb{I}(|\omega| \leq h^{-1}) d\omega + \frac{B^2}{4\pi^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2 + 1)^s} \mathbb{I}(|\omega| > h^{-1}) d\omega \quad (2.9)$$

where $\|g\|_{\infty} = \sup_x g(x)$. In order to analyze the right-hand side of (2.9) and similar expressions, the following statement is helpful.

Lemma 1 *Let $H(n, h) \equiv H(n, h; A_1, A_2, b, d, \beta, \gamma)$ where*

$$H(n, h) = h^{2A_1} \exp(-2dh^{-b}) + n^{-1} \int_0^{1/h} (\omega^2 + 1)^{A_2} \exp(2\gamma\omega^{\beta} - 2d\omega^b) d\omega. \quad (2.10)$$

Denote

$$\tilde{h}_n = \arg \min_h H(n, h), \quad \Delta_n \equiv \Delta_n(A_1, A_2, b, d, \beta, \gamma) = H(n, \tilde{h}_n; A_1, A_2, b, d, \beta, \gamma)$$

Then

$$\begin{array}{ll} \Delta_n \asymp n^{-1}, & \tilde{h}_n = 0 & \text{if } b > \beta, \\ \Delta_n \asymp n^{-1}, & \tilde{h}_n = 0 & \text{if } b = \beta, d > \gamma > 0 \\ \Delta_n \asymp n^{-1}, & \tilde{h}_n = 0 & \text{if } b = \beta, d = \gamma, A_2 < -1/2 \\ \Delta_n \asymp n^{-1} \log \log n, & \tilde{h}_n = [\log n / (2\gamma)]^{-1/\beta} & \text{if } b = \beta > 0, d = \gamma > 0, A_2 = -1/2 \\ \Delta_n \asymp n^{-1} (\log n)^{\frac{2A_2+1}{b}}, & \tilde{h}_n = [\log n / (2\gamma)]^{-1/\beta} & \text{if } b = \beta > 0, d = \gamma > 0, A_2 > -1/2 \\ \Delta_n \asymp n^{-1} \log n, & \tilde{h}_n = n^{-\frac{1}{2A_1}}, & \text{if } b = \beta = 0, d = \gamma = 0, A_2 = -1/2 \\ \Delta_n \asymp n^{-\frac{2A_1}{2A_1+2A_2+1}}, & \tilde{h}_n = n^{-\frac{1}{2A_1+2A_2+1}} & \text{if } b = \beta = 0, d = \gamma = 0, A_2 > -1/2 \\ \Delta_n \asymp (\log n)^{-V_1} n^{-d/\gamma}, & \tilde{h}_n = h^*(n) & \text{if } b = \beta > 0, \gamma > d > 0 \\ \Delta_n \asymp (\log n)^{-\frac{2A_1}{\beta}} \exp(-2d[h^*(n)]^{-b}), & \tilde{h}_n = h^*(n) & \text{if } \beta > b > 0, d > 0, \gamma > 0 \\ \Delta_n \asymp (\log n)^{-\frac{2A_1}{\beta}}, & \tilde{h}_n = h^*(n) & \text{if } b = d = 0, \beta > 0, \gamma > 0 \end{array} \quad (2.11)$$

where

$$h^*(n) = \left[\frac{1}{2\gamma} \left(\log n - \frac{2A_1 + 2A_2 + 1 - \beta}{\beta} \log \log n \right) \right]^{-\frac{1}{\beta}}, \quad V_1 = \frac{2A_1}{\beta} - \frac{d(2A_1 + 2A_2 + 1 - \beta)}{\beta \gamma}.$$

Validity of Lemma 1 can be verified by straightforward calculations. The upper bounds for the minimax risk follow directly from (2.11).

Theorem 1 *If g is bounded above, then, under Assumptions A1 and A2 (inequalities (2.2) and (2.5) only), one derives*

$$R_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B)) \leq C \Delta_n(s + a + (b - 1)/2, \alpha - a, b, d, \beta, \gamma) \quad (2.12)$$

where the expressions for \tilde{h}_n and $\Delta_n(A_1, A_2, b, d, \beta, \gamma)$ are given by (2.11).

Theorem 1 allows to derive upper bounds for the risk of $\widehat{\Phi}_{\tilde{h}_n}$ as an estimator of functional Φ_n defined in (1.3).

Corollary 1 *Let θ_i , $i = 1, \dots, n$, in (1.2) be i.i.d. with the pdf f . If $\varphi(\theta)$ is uniformly bounded $|\varphi(\theta)| \leq \|\varphi\|_\infty < \infty$, then under assumptions of Theorem 1, one has*

$$R_n(\widehat{\Phi}_{\tilde{h}_n}, \Phi_n, \Omega_s(B)) = \sup_{f \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_{\tilde{h}_n} - \Phi_n)^2 \leq 2R_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B)) + 2n^{-1}\|\varphi\|_\infty^2 \leq CR_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B)), \quad (2.13)$$

where $R_n(\widehat{\Phi}_{\tilde{h}_n}, \Omega_s(B))$ is given by expression (2.12) and $C > 0$ is a constant independent of n .

2.3 Adaptive estimation

Note that in the expressions for the optimal value of bandwidth \tilde{h}_n in Lemma 1, parameters a, α, b, d, β and γ are known, the only unknown parameters are s and B . Hence, the only cases when one needs an adaptive choice of bandwidth are the cases where \tilde{h}_n depends on A_1 . However, in every situation except the case when $d = b = \gamma = \beta = 0$ and $a < \alpha + 1/2$, one can easily find an alternative value \hat{h}_n of h that delivers a nearly optimal (up to at most a logarithmic factor) convergence rates.

In the case when $d = b = \gamma = \beta = 0$ and $a < \alpha + 1/2$, one can use the Lepskii method for construction of \hat{h}_n (see, e.g., Lepski (1991), Lepski *et al.* (1997)). In order to apply the method, consider the set of bandwidths

$$\mathcal{H} = \left\{ h_j = n^{-\frac{1}{2\alpha}} 2^{-j}, j = 0, 1, \dots, J \right\} \quad \text{with} \quad 2^J \leq (\log n)^{-1} n^{\frac{2\alpha-1}{2\alpha}}, \quad (2.14)$$

where J is the largest positive integer satisfying inequality above. Denote

$$\hat{j} = \min \left\{ j : 0 \leq j \leq J; |\widehat{\Phi}_{h_j} - \widehat{\Phi}_{h_k}| \leq C_\Phi n^{-1/2} \sqrt{\log n} h_k^{-(\alpha-a+1/2)}, \quad \forall k = j, \dots, J \right\}, \quad (2.15)$$

where C_Φ is such that

$$C_\Phi \geq 4 \max \left\{ \frac{C_{\varphi 2}}{\pi \sqrt{\log n}}; \frac{C_{\varphi 2}}{C_{g1}} \left(\frac{16}{3\pi} + \frac{2\sqrt{\|g\|_\infty}}{\sqrt{\pi}} \right) \right\} \quad (2.16)$$

and constants $C_{\varphi 2}$ and C_{g1} appear in (2.5) and (2.2), respectively. The following statement provides the minimax upper bounds for the risk when the bandwidth h_n is chosen adaptively, without the knowledge of parameters B and s .

Theorem 2 *If g is bounded above, then, under Assumptions A1 and A2 (inequalities (2.2) and (2.5) only), one obtains the following expressions for $\widehat{R}_n \equiv R_n(\widehat{\Phi}_{\widehat{h}}, \Omega_s(B))$*

$$\begin{aligned}
\widehat{R}_n &\asymp n^{-1}, \widehat{h}_n = 0 & \text{if } b > \beta, \\
\widehat{R}_n &\asymp n^{-1}, \widehat{h}_n = 0 & \text{if } b = \beta, d > \gamma > 0 \\
\widehat{R}_n &\asymp n^{-1}, \widehat{h}_n = 0 & \text{if } b = \beta, d = \gamma, a > \alpha + 1/2 \\
\widehat{R}_n &\asymp n^{-1} \log \log n, \widehat{h}_n = \widehat{h}_n^* & \text{if } b = \beta > 0, d = \gamma > 0, a = \alpha + 1/2 \\
\widehat{R}_n &\asymp n^{-1} \log n, \widehat{h}_n = n^{-\frac{1}{2a-1}} & \text{if } b = \beta = 0, d = \gamma = 0, a = \alpha + 1/2 \\
\widehat{R}_n &\asymp n^{-1} (\log n)^{\frac{2\alpha-2a+1}{\beta}}, \widehat{h}_n = \widehat{h}_n^* & \text{if } b = \beta > 0, d = \gamma > 0, a < \alpha + 1/2 \\
\widehat{R}_n &\asymp n^{-\frac{2s+2a-1}{2s+2\alpha}} \log n, \widehat{h}_n = h_{\widehat{j}} & \text{if } b = \beta = 0, d = \gamma = 0, a < \alpha + 1/2 \\
\widehat{R}_n &\asymp (\log n)^{-\frac{U_1}{\beta}} n^{-d/\gamma}, \widehat{h}_n = \widehat{h}_n^* & \text{if } b = \beta > 0, \gamma > d > 0 \\
\widehat{R}_n &\asymp (\log n)^{-\frac{U_2}{\beta}} \exp\left(-2d \left[\frac{\log n}{2\gamma}\right]^{b/\beta}\right), \widehat{h}_n = \widehat{h}_n^* & \text{if } \beta > b > 0, d > 0, \gamma > 0, \\
\widehat{R}_n &\asymp (\log n)^{-\frac{2s+2a-1}{\beta}}, \widehat{h}_n = \widehat{h}_n^{**} & \text{if } b = d = 0, \beta > 0, \gamma > 0
\end{aligned} \tag{2.17}$$

Here, $\widehat{h}_n^* = [\log n / (2\gamma)]^{-\frac{1}{\beta}}$, $\widehat{h}_n^{**} = [\log n / (3\gamma)]^{-\frac{1}{\beta}}$, \widehat{j} is defined in (2.15) with C_Φ defined in (2.16), $U_1 = \min(\beta + 2a + 2s - 1, \beta + 2a - 2\alpha - 1)$ and $U_2 = b + 2a + 2s - 1$.

Careful comparisons between the rates in Theorems 1 and 2 reveal that the rates coincide up to a constant except for the cases when $b = \beta = 0, d = \gamma = 0, a < \alpha + 1/2$, or $b = \beta > 0, \gamma > d > 0$, or $\beta > b > 0, d > 0, \gamma > 0$; in the latter cases the rates coincide up to a log-factor of n . One can also replace \widehat{h}_n by \widehat{h}_n in Corollary 1 and obtain adaptive minimax upper bounds for the risk of the estimator of Φ_n .

Remark 1 (Comparison with Butucea and Comte (2009)). Note that, since we are looking at the case of Sobolev classes only, we can examine the rate of convergence of the non-specific term v_n used in Butucea and Comte (2009) and obtain exact results. In particular, in terms of rates of convergence, there are three parametric regions ($b > \beta$), or ($b = \beta, d > \gamma$), or ($b = \beta, d = \gamma, a > \alpha + 1/2$), several nearly parametric regions ($b = \beta, d = \gamma, a \leq \alpha + 1/2$), two different regions of polynomial rates of convergence ($b = \beta, \gamma > d > 0$ or $d = \gamma = 0, a \leq \alpha + 1/2$), a region of logarithmic rates ($b = d = 0, \beta > 0, \gamma > 0$) and an “in-between” region ($\beta > b > 0, d > 0, \gamma > 0$) where $R_n(\widehat{\Phi}_{\widehat{h}_n}, \Omega(B))$ converges to zero faster than $(\log n)^{-B_1}$ and slower than n^{-B_2} for any $B_1 > 0, B_2 > 0$.

2.4 The lower bounds for the risk

Theorem 1 provides upper bounds for the risk for any combination of parameters $s, a, b, d, \alpha, \beta$ and γ . However, to the best of our knowledge, the minimax lower bounds for the risk of a general linear functional have not been obtained so far. Butucea and Comte (2009) derived those lower bounds only in the simple case when $|\varphi^*(\omega)| = 1$. Below, we derive the minimax lower bounds for the risk and show that, for a wide range of functions φ , the upper bounds (2.12) match the lower bounds up to a constant or a logarithmic factor of n .

Denote

$$R_n(\Omega_s(B)) = \inf_{\Phi} \sup_{f \in \Omega_s(B)} \mathbb{E}(\widetilde{\Phi} - \Phi)^2 \tag{2.18}$$

where $\tilde{\Phi}$ is any estimator of Φ based on observations Y_1, \dots, Y_n . Then, the following theorem is true.

Theorem 3 *Let g be bounded above and such that function g^* is differentiable and*

$$\frac{|(g^*)'(\omega)|}{|g^*(\omega)|} \leq C_g (1 + |\omega|)^\tau, \quad \tau \geq 0, \quad \text{with } \tau = 0 \quad \text{if } \gamma = 0. \quad (2.19)$$

Let there exist $\omega_0 \in (0, \infty)$ such that, for $|\omega| > \omega_0$, function $\rho(\omega) = \arg(\varphi^(\omega))$ is twice continuously differentiable with $|\rho^{(j)}(\omega)| \leq \rho < \infty$, $j = 0, 1, 2$. Then, under Assumptions A1 and A2 (inequalities (2.3) and (2.4) only), one derives*

$$R_n(\Omega_s(B)) \geq \begin{cases} C n^{-1} & \text{if } b > \beta \text{ or } b = \beta, d > \gamma > 0, \\ C n^{-1} & \text{if } b = \beta, d = \gamma, a \geq \alpha + 1/2, \\ C n^{-1} & \text{if } b = \beta > 0, d = \gamma > 0, a < \alpha + 1/2, \\ C n^{-\frac{2s+2a-1}{2s+2\alpha}} & \text{if } d = \gamma = 0, a < \alpha + 1/2 \\ C (\log n)^{-\frac{U_3}{\beta}} n^{-d/\gamma} & \text{if } b = \beta, \gamma \geq d > 0 \\ C (\log n)^{-\frac{U_4}{\beta}} \exp\left(-2d \left[\frac{\log n}{2\gamma}\right]^{b/\beta}\right) & \text{if } b < \beta, d > 0, \gamma > 0, \\ C (\log n)^{-\frac{2s+2a-1}{\beta}} & \text{if } b = d = 0, \beta > 0, \gamma > 0 \end{cases} \quad (2.20)$$

where $U_3 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 8\beta - (U_\tau + 1)d/\gamma - 1$, $U_4 = 2a + 2s + 8b - 1$ and $U_\tau = \min(7\beta - 2\tau - 1, 5\beta + 1)$.

Observe that for in the cases when $\beta < b$, or $\beta = b, d > \gamma > 0$ or $\beta = b, d = \gamma, a \geq \alpha + 1/2$, the rates of convergence in Theorem 1 are parametric and, hence, cannot be improved. In addition, the lower and upper bounds coincide up to a constant if $d = \gamma = 0$, $a < \alpha + 1/2$ or $b = d = 0$, $\beta > 0, \gamma > 0$; otherwise, they coincide up to a logarithmic factor.

Since any estimator $\tilde{\Phi}_n$ of Φ_n defined in (1.3) can be viewed as an estimator of Φ , due to inequality

$$\mathbb{E}(\tilde{\Phi}_n - \Phi_n)^2 \geq 0.5 \mathbb{E}(\tilde{\Phi}_n - \Phi)^2 - 2n^{-1} \|\varphi\|_\infty^2,$$

Theorem 3 immediately provides the lower bounds for the risk of any estimator $\tilde{\Phi}_n$ of Φ_n based on Y_1, \dots, Y_n .

Corollary 2 *Let θ_i , $i = 1, \dots, n$, in (1.2) be i.i.d. with pdf f . If $\varphi(\theta)$ is uniformly bounded $|\varphi(\theta)| \leq \|\varphi\|_\infty < \infty$, then under assumptions of Theorem 3, for sufficiently large n , one has*

$$R_n(\Phi_n, \Omega_s(B)) = \inf_{\tilde{\Phi}_n} \sup_{f \in \Omega_s(B)} \mathbb{E}(\tilde{\Phi}_n - \Phi_n)^2 \geq C R_n(\Omega_s(B)), \quad (2.21)$$

where $R_n(\Omega_s(B))$ is given by expression (2.20) and $C > 0$ is a constant independent of n .

As an example of application of theory above, we solve the problem of pointwise estimation of the mixing density with classical and Berkson errors studied by Delaigle (2007).

2.5 Pointwise estimation of the deconvolution density with classical and Berkson errors

Consider the situation where one is interested in estimating the pdf f_ζ of the random variable $\zeta = \theta + \eta$ where θ and η are independent, the pdf f_η of η is known and one has measurements Y_1, \dots, Y_n of random variable $Y = \theta + \xi$ of the form of (1.2) where the pdf g of ξ is known. The model was originally introduced by Berkson (1950) in the regression context and subsequently studied by Delaigle (2007) who obtained the upper bounds for the integrated mean squared risk. In particular, if the pdf $q(y)$ of Y is k times continuously differentiable and is such that q and $q^{(k+1)}$ are square integrable and $q^{(k+1)}$ is bounded, Delaigle (2007) derived

$$\mathbb{E}\|\hat{f}_\zeta - f_\zeta\|^2 \leq \begin{cases} Cn^{-(2k)/(2k+1-2\beta)}, & \text{if } |f_\eta^*(\omega)/g^*(\omega)| \asymp |\omega|^{-\beta} \text{ for } |\omega| \rightarrow \infty, \\ C(\log n)^{-2k/\beta}, & \text{if } |f_\eta^*(\omega)/g^*(\omega)| \asymp |\omega|^b \exp(\gamma|\omega|^\beta) \text{ for } |\omega| \rightarrow \infty, \end{cases}$$

where $\|\cdot\|$ denotes the L^2 -norm with respect to the Lebesgue measure and the constant C depends on the density f of each θ .

The theory developed in this paper allows one to construct an estimator of the pdf f_ζ at a point x_0 with no additional effort. Let, as before, f , g and q be the pdfs of θ , ξ and Y , respectively. Then $f_\zeta^* = f^* f_\eta^* = q^* f_\eta^* / g^*$ and

$$f_\zeta(x_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix_0\omega} \frac{q^*(\omega)f_\eta^*(\omega)}{g^*(\omega)} d\omega,$$

Therefore, $\varphi^*(\omega) = e^{ix_0\omega} f_\eta^*(-\omega)$, so that $|\varphi^*(\omega)| = |f_\eta^*(\omega)|$. The estimator of $f_\zeta(x_0)$ is of the form (2.7) and Theorems 1 and 3 give the upper and the minimax lower bounds for the risk of estimating f_ζ at a point x_0 . In addition, Theorem 2 provides an adaptive estimator of $f_\zeta(x_0)$ that, to the best of our knowledge, has not been derived so far.

Here, one can observe an interesting phenomenon that we obtain a wider variety of convergence rates here than Delaigle (2007) who recovered only parametric, polynomial (with $d = 0$) and logarithmic convergence rates. The more diverse convergence rates in our case are not due to the fact that we are studying local (pointwise) error while Delaigle (2007) was interested in the global one. Indeed, for this particular example, with a little effort, our minimax theory can be extended to the situation of the global error. The reason for the wider diversity lies in the fact that we impose assumptions on g^* and f_η^* separately while Delaigle (2007) considers only the cases when the absolute value of the ratio $|f_\eta^*/g^*|$ grows polynomially or exponentially as $|\omega| \rightarrow \infty$.

3 Estimation of linear functionals by using inversion formulas

3.1 Formulation and some inversion formulae

Estimation of the linear functional Φ in (1.1) relies on the fact that $\varphi \in L^2(-\infty, \infty)$, so its Fourier transform exists. It is easy to see that this condition, however, is not necessary for consistent estimation of Φ . Consider, for example, estimation of the m -th moment of $f(\theta)$

$$\Phi_m = \int_{-\infty}^{\infty} \theta^m f(\theta) d\theta. \quad (3.1)$$

Note that if $\psi_m(y)$ is a solution of the equation

$$\int_{-\infty}^{\infty} g(y - \theta) \psi_m(y) dy = \theta^m \quad (3.2)$$

then $\Phi_m = \int_{-\infty}^{\infty} \psi_m(y) q(y) dy = \mathbb{E}[\psi_m(Y)]$. In order to construct $\psi_m(y)$ satisfying equation (3.2), denote

$$\mu_k = \int_{-\infty}^{\infty} \theta^k g(\theta) d\theta, \quad \nu_k = \int_{-\infty}^{\infty} \theta^k f(\theta) d\theta, \quad (3.3)$$

and assume that $\mu_{2m} < \infty$ and $\nu_{2m} < \infty$. Let $c_m = 1$ and $c_k, k = 0, \dots, m-1$, be solutions of the system of linear equations

$$\sum_{k=j}^m \mu_{k-j} c_k = 0, \quad j = 0, 1, \dots, m-1.$$

Then, it is easy to check that

$$\psi_m(y) = y^m + \sum_{k=0}^{m-1} c_k y^k,$$

and, under assumption that $\mu_{2m} < \infty$ and $\nu_{2m} < \infty$, Φ_m can be estimated by

$$\hat{\Phi}_m = n^{-1} \sum_{l=1}^n \psi_m(Y_l), \quad \text{where} \quad \mathbb{E}\hat{\Phi}_m = \Phi_m, \quad \text{Var}[\hat{\Phi}_m] \leq C_m n^{-1} \quad (3.4)$$

and constant C_m depends only on m, μ_{2m} and ν_{2m} .

Note that although we did not use Fourier transform for estimation of Φ_m and Fourier transform of x^m does not exist in a regular sense, it does exist in a sense of generalized functions and is equal to $(-1)^m \delta^{(m)}(\omega)$ where $\delta^{(m)}(\omega)$ is the m -th derivative of the Dirac delta function (see, e.g. Zayed (1996)). However, using the Fourier transform of φ as a generalized function would require f to belong to a so called test-function space. Those spaces are usually very restrictive, like, e.g., commonly used for the Fourier transforms of generalized functions, the space of the Schwartz distributions which consists of all infinitely differentiable functions that vanish outside some compact set (see, e.g. Zayed (1996)). One, of course, cannot expect the unknown density f to belong to such space and, moreover, this will make any minimax estimation totally irrelevant. For this reason, instead of using the theory of generalized functions we shall use inversion formulas that mimic generalized functions but do not require unreasonable assumptions on the unknown pdf f . Our goal is to represent the functionals of interest as integrals of the Fourier transform of f^* and its derivatives. Dattner *et al.* (2011) used inversion formula of Gil-Pelaez (1951) for estimation of the cumulative distribution function at a point, nevertheless, there are many more possible applications of this technique. Below we consider several examples.

Example 1. Pointwise estimation of the deconvolution cumulative distribution function. In order to represent the cdf $F(t)$ of θ at a point t we apply formula 3.721.1 of Gradshteyn and Ryzhik (1980)

$$\int_0^{\infty} \frac{\sin(ax)}{x} dx = \text{sign}(a), \quad (3.5)$$

where $\text{sign}(a)$ is the sign of a . Denote the real and the imaginary part of z by $\Re[z]$ and $\Im[z]$, respectively. Then, due to the relation $\mathbb{I}(\theta \leq t) = 1/2 - 1/2 \text{sign}(\theta - t)$, the cdf $F(t)$ can be

represented as $F(t) = 0.5 - 0.5 \Phi(t)$ where

$$\begin{aligned}
\Phi(t) &= \int_{-\infty}^{\infty} \text{sign}(\theta - t) f(\theta) d\theta = \frac{2}{\pi} \int_{-\infty}^{\infty} \left[\int_0^{\infty} \frac{\sin((\theta - t)\omega)}{\omega} d\omega \right] f(\theta) d\theta \\
&= \frac{2}{\pi} \int_0^{\infty} \frac{\cos(t\omega)}{\omega} \left[\int_{-\infty}^{\infty} \sin(\theta\omega) f(\theta) d\theta \right] d\omega - \frac{2}{\pi} \int_0^{\infty} \frac{\sin(t\omega)}{\omega} \left[\int_{-\infty}^{\infty} \cos(\theta\omega) f(\theta) d\theta \right] d\omega \\
&= \frac{2}{\pi} \int_0^{\infty} \frac{\cos(t\omega)}{\omega} \Im[f^*(\omega)] d\omega - \frac{2}{\pi} \int_0^{\infty} \frac{\sin(t\omega)}{\omega} \Re[f^*(\omega)] d\omega.
\end{aligned} \tag{3.6}$$

Example 2. Estimation of truncated moments. Consider estimation of Φ of the form (1.1) where $\varphi(\theta) = \theta^m \mathbb{I}(\theta > \theta_0)$ or $\varphi(\theta) = \theta^m \text{sign}(\theta - \theta_0)$. It is easy to see that these two functionals are related as

$$\int_{-\infty}^{\infty} \theta^m \mathbb{I}(\theta > \theta_0) f(\theta) d\theta = \frac{1}{2} \Phi_m + \frac{1}{2} \int_{-\infty}^{\infty} \theta^m \text{sign}(\theta - \theta_0) f(\theta) d\theta$$

where Φ_m is the m -th moment of f defined in (3.1). Since Φ_m can be estimated by $\hat{\Phi}_m$ defined in (3.4) with parametric convergence rates, it is sufficient to consider estimation of

$$\Phi(m, \theta_0) = \int_{-\infty}^{\infty} \theta^m \text{sign}(\theta - \theta_0) f(\theta) d\theta. \tag{3.7}$$

Using formula (3.5), rewrite $\Phi(m, \theta_0)$ as

$$\Phi(m, \theta_0) = \frac{2}{\pi} \int_0^{\infty} \frac{\cos(\omega\theta_0)}{\omega} \int_{-\infty}^{\infty} \sin(\omega\theta) \theta^m f(\theta) d\theta - \frac{2}{\pi} \int_0^{\infty} \frac{\sin(\omega\theta_0)}{\omega} \int_{-\infty}^{\infty} \cos(\omega\theta) \theta^m f(\theta) d\theta.$$

Observe that

$$\int_{-\infty}^{\infty} \sin(\omega\theta) \theta^m f(\theta) d\theta = i^{-m} \Im \left[\frac{d^m}{d\omega^m} \int_{-\infty}^{\infty} e^{i\omega\theta} f(\theta) d\theta \right] = \begin{cases} (-1)^k \frac{d^{2k}}{d\omega^{2k}} \Im[f^*(\omega)], & m = 2k, \\ (-1)^{k+1} \frac{d^{2k+1}}{d\omega^{2k+1}} \Re[f^*(\omega)], & m = 2k + 1, \end{cases}$$

and, similarly,

$$\int_{-\infty}^{\infty} \cos(\omega\theta) \theta^m f(\theta) d\theta = \begin{cases} (-1)^k \frac{d^{2k}}{d\omega^{2k}} \Re[f^*(\omega)], & m = 2k, \\ (-1)^k \frac{d^{2k+1}}{d\omega^{2k+1}} \Im[f^*(\omega)], & m = 2k + 1. \end{cases}$$

Combining both cases, we obtain

$$\Phi(m, \theta_0) = \begin{cases} \frac{(-1)^k 2}{\pi} \left[\int_0^{\infty} \frac{\cos(\omega\theta_0)}{\omega} \frac{d^{2k} \Im[f^*(\omega)]}{d\omega^{2k}} d\omega - \int_0^{\infty} \frac{\sin(\omega\theta_0)}{\omega} \frac{d^{2k+1} \Re[f^*(\omega)]}{d\omega^{2k+1}} d\omega \right], & m = 2k, \\ \frac{(-1)^{k+1} 2}{\pi} \left[\int_0^{\infty} \frac{\cos(\omega\theta_0)}{\omega} \frac{d^{2k+1} \Re[f^*(\omega)]}{d\omega^{2k+1}} d\omega + \int_0^{\infty} \frac{\sin(\omega\theta_0)}{\omega} \frac{d^{2k+1} \Im[f^*(\omega)]}{d\omega^{2k+1}} d\omega \right], & m = 2k + 1. \end{cases} \tag{3.8}$$

Example 3. Estimation of generalized moments. Consider estimation of functionals of the form (1.1) where $\varphi(\theta) = \theta^m u(\theta)$ with $u(\theta)$ is such that $u(\theta) \in L^2(-\infty, \infty)$ but $\theta^m u(\theta) \notin L^2(-\infty, \infty)$. Note that, since $u(\theta) \in L^2(-\infty, \infty)$ implies, for any θ_0 , that $u(\theta)\mathbb{I}(\theta > \theta_0) \in L^2(-\infty, \infty)$ and $u(\theta)\text{sign}(\theta - \theta_0) \in L^2(-\infty, \infty)$, we are automatically including all square integrable discontinuous functions $u(\theta)$. On the other hand, since functions $\mathbb{I}(\theta > \theta_0)$ and $\text{sign}(\theta - \theta_0)$ are not square integrable, Example 2 is not a particular case of Example 3.

In order to derive an inversion formula for this functional, denote $f_m(\theta) = \theta^m f(\theta)$ and observe that $f_m^*(\omega) = i^{-m} \frac{d^m}{d\omega^m} [f^*(\omega)]$. Therefore,

$$\Phi_u = \int_{-\infty}^{\infty} \theta^m u(\theta) f(\theta) d\theta = \frac{i^{-m}}{2\pi} \int_{-\infty}^{\infty} u^*(-\omega) \frac{d^m f^*(\omega)}{d\omega^m} d\omega. \quad (3.9)$$

3.2 Construction of the estimators and evaluation of their risks

Observe that in all three examples above, the linear functionals Φ can be presented as a combination of two integrals

$$\Phi = \int_0^{\infty} \psi_{m1}^*(\omega) \frac{d^m \Re[f^*(\omega)]}{d\omega^m} d\omega + \int_0^{\infty} \psi_{m2}^*(\omega) \frac{d^m \Im[f^*(\omega)]}{d\omega^m} d\omega. \quad (3.10)$$

where we assume that f satisfies conditions (that, of course, depend on the particular forms of $\psi_{m1}^*(\omega)$ and $\psi_{m2}^*(\omega)$) which guarantee absolute convergence of the integrals in (3.10). In particular, we assume that both $f^*(\omega)$ and $g^*(\omega)$ are m times differentiable. Note that regular case corresponds to $m = 0$ and $\psi_{m1}^*(\omega) = \psi_{m2}^*(\omega) = \varphi^*(-\omega)/\pi$.

In order to construct an estimator of the functional Φ in (3.10), we partition the area of integration into $\mathcal{A}_1 = [0; 1]$ and $\mathcal{A}_2 = (1, \infty)$ and rewrite Φ as $\Phi = \Phi_1 + \Phi_2$ where

$$\Phi_k = \int_{\mathcal{A}_k} \psi_{m1}^*(\omega) \Re \left[\frac{d^m f^*(\omega)}{d\omega^m} \right] d\omega + \int_{\mathcal{A}_k} \psi_{m2}^*(\omega) \Im \left[\frac{d^m f^*(\omega)}{d\omega^m} \right] d\omega, \quad k = 1, 2. \quad (3.11)$$

Without loss of generality, we assume that g is an even function, so that its Fourier transform $g^*(\omega)$ is real-valued. One can easily extend our study to the case when g is an arbitrary pdf; we leave this case for the reader to examine. Note that using the general Leibniz rule, Φ_1 in (3.11) can be re-written as

$$\Phi_1 = \sum_{j=0}^m \binom{m}{j} \int_0^1 \frac{d^{(m-j)}}{d\omega^{(m-j)}} \left[\frac{1}{g^*(\omega)} \right] [\psi_{m1}^*(\omega) u_{j1}(\omega) + \psi_{m2}^*(\omega) u_{j2}(\omega)] d\omega \quad (3.12)$$

where

$$u_{j1}(\omega) = \Re \left[\frac{d^j q^*(\omega)}{d\omega^j} \right], \quad u_{j2}(\omega) = \Im \left[\frac{d^j q^*(\omega)}{d\omega^j} \right] \quad (3.13)$$

Denote

$$v_{j1}(\omega) = \int_{-\infty}^{\infty} y^j q(y) \cos(\omega y) dy, \quad v_{j2}(\omega) = \int_{-\infty}^{\infty} y^j q(y) \sin(\omega y) dy$$

and construct their respective unbiased estimators as

$$\widehat{v}_{j1}(\omega) = n^{-1} \sum_{l=1}^n Y_l^j \cos(\omega Y_l), \quad \widehat{v}_{j2}(\omega) = n^{-1} \sum_{l=1}^n Y_l^j \sin(\omega Y_l). \quad (3.14)$$

By taking derivatives of $q^*(\omega)$ under the integral sign, it is easy to check that the unbiased estimators of $u_{j1}(\omega)$ and $u_{j2}(\omega)$ are, respectively, given by

$$\widehat{u}_{j1}(\omega) = \begin{cases} (-1)^{j/2} \widehat{v}_{j1}(\omega), & \text{if } j \text{ is even} \\ (-1)^{(j+1)/2} \widehat{v}_{j2}(\omega), & \text{if } j \text{ is odd} \end{cases} \quad (3.15)$$

$$\widehat{u}_{j2}(\omega) = \begin{cases} (-1)^{j/2} \widehat{v}_{j2}(\omega), & \text{if } j \text{ is even} \\ (-1)^{(j-1)/2} \widehat{v}_{j1}(\omega), & \text{if } j \text{ is odd} \end{cases} \quad (3.16)$$

Combination of formulae (3.12) – (3.16) imply that Φ_1 can be estimated by

$$\widehat{\Phi}_1 = \sum_{j=0}^m \binom{m}{j} \int_0^1 \frac{d^{(m-j)}}{d\omega^{(m-j)}} \left[\frac{1}{g^*(\omega)} \right] [\psi_{m1}^*(\omega) \widehat{u}_{j1}(\omega) + \psi_{m2}^*(\omega) \widehat{u}_{j2}(\omega)] d\omega. \quad (3.17)$$

In order to estimate Φ_2 , using integration by parts, partition Φ_2 in (3.11) as $\Phi_2 = F_m(1) + \Phi_{20}$, where

$$\begin{aligned} F_m(1) &= \sum_{k=1}^m (-1)^k \left[\frac{d^{k-1} \psi_{m1}^*(\omega)}{d\omega^{k-1}} \frac{d^{m-k} [\Re(f^*(\omega))]}{d\omega^{m-k}} + \frac{d^{k-1} \psi_{m2}^*(\omega)}{d\omega^{k-1}} \frac{d^{m-k} [\Im(f^*(\omega))]}{d\omega^{m-k}} \right]_{\omega=1}, \\ \Phi_{20} &= (-1)^m \int_1^\infty \left(\frac{d^m \psi_{m1}^*(\omega)}{d\omega^m} \Re[f^*(\omega)] + \frac{d^m \psi_{m2}^*(\omega)}{d\omega^m} \Im[f^*(\omega)] \right) d\omega. \end{aligned}$$

Again, taking into account that $f^*(\omega) = q^*(\omega)/g^*(\omega)$ and applying the general Leibniz rule, rewrite $F_m(1)$ and Φ_{20} as

$$\begin{aligned} F_m(1) &= \sum_{k=1}^m \sum_{j=0}^{m-k} (-1)^k \binom{m-k}{j} [A_{m,j,k,1}(1) u_{j1}(1) + A_{m,j,k,2}(1) u_{j2}(1)], \\ \Phi_{20} &= (-1)^m \int_1^\infty \left[\frac{d^m \psi_{m1}^*(\omega)}{d\omega^m} u_{01}(\omega) + \frac{d^m \psi_{m2}^*(\omega)}{d\omega^m} u_{02}(\omega) \right] \frac{1}{g^*(\omega)} d\omega, \end{aligned}$$

where $u_{j1}(\omega)$ and $u_{j2}(\omega)$ are defined in (3.13) and

$$A_{m,j,k,l}(\omega) = \frac{d^{k-1} \psi_{lm}^*(\omega)}{d\omega^{k-1}} \frac{d^{m-k-j}}{d\omega^{m-k-j}} \left(\frac{1}{g^*(\omega)} \right), \quad l = 1, 2. \quad (3.18)$$

Therefore, we can estimate Φ_2 by $\widehat{\Phi}_{2h} = \widehat{F}_m(1) + \widehat{\Phi}_{20h}$ where

$$\widehat{F}_m(1) = \sum_{k=1}^m \sum_{j=0}^{m-k} (-1)^k \binom{m-k}{j} [A_{m,j,k,1}(1) \widehat{u}_{j1}(1) + A_{m,j,k,2}(1) \widehat{u}_{j2}(1)] \quad (3.19)$$

$$\widehat{\Phi}_{20h} = (-1)^m \int_1^{1/h} \left[\frac{d^m \psi_{m1}^*(\omega)}{d\omega^m} \widehat{u}_{01}(\omega) + \frac{d^m \psi_{m2}^*(\omega)}{d\omega^m} \widehat{u}_{02}(\omega) \right] \frac{1}{g^*(\omega)} d\omega, \quad (3.20)$$

and $\widehat{u}_{j1}(\omega)$ and $\widehat{u}_{j2}(\omega)$ are defined by (3.15) and (3.16). Finally, we estimate Φ in (3.10) by $\widehat{\Phi}_h = \widehat{\Phi}_1 + \widehat{F}_m(1) + \widehat{\Phi}_{20h}$ where $\widehat{\Phi}_1$, $\widehat{F}_m(1)$ and $\widehat{\Phi}_{20h}$ are evaluated according to (3.17), (3.19) and (3.20), respectively.

In order to construct an upper bound for the risk of the estimator $\widehat{\Phi}_h$, we denote

$$\sigma_{j1}^2(\omega) = n \text{Var}(\widehat{u}_{j1}), \quad \sigma_{j2}^2(\omega) = n \text{Var}(\widehat{u}_{j2}) \quad (3.21)$$

and consider a class of pdfs

$$\Xi_s(B) = \left\{ f : \sup_{\omega} [|f^*(\omega)| (|\omega|^s + 1)] \leq B_2 \right\} \quad (3.22)$$

Then, the risk of the estimator $\widehat{\Phi}_h$ is given by the following statement.

Theorem 4 *Assume that functions f and g are such that*

$$\mu_{2m} < \infty, \quad \nu_{2m} < \infty, \quad (3.23)$$

where μ_k and ν_k are defined in (3.3). Let also function g^* be real-valued, satisfy Assumption A1, be m times differentiable and such that, for some $C_g > 0$

$$\left| \frac{1}{g^*(\omega)} \frac{d^j g^*(\omega)}{d\omega^j} \right| \leq C_g (|\omega| + 1)^{j\tau}, \quad \tau \geq 0, j = 0, \dots, m, \quad \text{where } \tau = 0 \text{ if } \gamma = 0. \quad (3.24)$$

Let $\psi_{m1}^*(\omega)$ and $\psi_{m2}^*(\omega)$ be such that for some positive C_ψ and nonnegative d, a_m and b , and for $|\omega| \geq 1$

$$\left| \frac{d^j \psi_{mk}^*(\omega)}{d\omega^j} \right| \leq C_\psi (\omega^2 + 1)^{-a_m(j+1)/2} \exp(-d|\omega|^b), \quad j = 0, \dots, m, \quad k = 1, 2. \quad (3.25)$$

Assume also that there exists an absolute constant C_σ such that

$$\int_0^1 \left[\frac{d^{m-j}}{d\omega^{m-j}} \left(\frac{1}{g^*(\omega)} \right) \right]^2 |\psi_{mk}^*(\omega)|^2 \sigma_{jk}^2(\omega) d\omega \leq C_\sigma, \quad j = 0, \dots, m, \quad . \quad (3.26)$$

Let $\widehat{\Phi}_h = \widehat{\Phi}_1 + \widehat{F}_m(1) + \widehat{\Phi}_{20h}$ where $\widehat{\Phi}_1$, $\widehat{F}_m(1)$ and $\widehat{\Phi}_{20h}$ are defined in (3.17), (3.19) and (3.20), respectively. Then,

$$\mathbb{E}(\widehat{\Phi}_h - \Phi)^2 \leq C \left[h^{2A} \exp(-2dh^{-b}) + n^{-1} \int_1^{h^{-1}} (\omega^2 + 1)^{\alpha-(m+1)a_m} \exp(-2d\omega^b + 2\gamma\omega^\beta) d\omega \right], \quad (3.27)$$

where $A = (m+1)a_m + s + b - 1$ if $f \in \Xi_s(B)$ and $A = (m+1)a_m + s + (b-1)/2$ if $f \in \Omega_s(B)$. Here, $\Omega_s(B)$ and $\Xi_s(B)$ are defined by (2.1) and (3.22), respectively.

Depending on the respective values of $b, d, \beta, \gamma, s, m$ and a_m , one can obtain convergence rates and optimal bandwidth values \tilde{h} for each combination of parameters using Lemma 1 in Section 2.2. Moreover, application of an equivalent of Theorem 2 allows one to obtain an adaptive estimator of Φ . Note, however, that convergence rates in Theorem 4 are not minimax. Indeed, in addition to $f \in \Omega_s(B)$ or $f \in \Xi_s(B)$, assumption (3.26) imposes additional conditions on f^* that depend, in a non-trivial way, on the shape of functions ψ_{mk}^* , $k = 1, 2$, thus, modifying the class of functions f . For this reason, one has to derive upper and lower bounds for the minimax risk on a case-by-case basis. We shall consider some examples in the next section.

3.3 Examples of estimation of linear functionals using inversion formulas

Example 1 (continuation). Pointwise estimation of the deconvolution cumulative distribution function. Recall that $F(t) = 0.5 - 0.5\Phi(t)$ where $\Phi(t)$ is defined by formula (3.6). Since $|\sin(x)/x| \leq 1$ for any x , one can easily show that both integrals in (3.6) are absolutely

convergent provided $\int_{-\infty}^{\infty} |\theta| f(\theta) d\theta < \infty$. Therefore, (3.6) is a particular case of (3.10) with $m = 0$, $\psi_{m1}^*(\omega) = -2 \sin(\omega)/(\pi\omega)$ and $\psi_{m2}^*(\omega) = 2 \cos(\omega)/(\pi\omega)$, so that inequality (3.24) is valid. Observe also that (3.25) holds with $m = 0$, $a_0 = 1$ and $b = d = 0$. Using notations above, one can write

$$\Phi(t) = \frac{2}{\pi} \int_0^\infty \frac{\cos(t\omega) v_{02}(\omega) - \sin(t\omega) v_{01}(\omega)}{\omega g^*(\omega)} d\omega,$$

so that

$$\widehat{\Phi}_h(t) = \frac{2}{\pi} \int_0^{h^{-1}} \frac{\cos(t\omega) \widehat{v}_{02}(\omega) - \sin(t\omega) \widehat{v}_{01}(\omega)}{\omega g^*(\omega)} d\omega.$$

Let g satisfy condition (2.2) and $\mu_2 < \infty$, $\nu_2 < \infty$, where μ_k and ν_k are defined in (3.3). Then, due to

$$\sigma_{01}^2(\omega) = \int_{-\infty}^{\infty} \cos(\omega y) q(y) dy \leq 1, \quad \sigma_{02}^2 = \int_{-\infty}^{\infty} \sin(\omega y) q(y) dy \leq \min(1, 2\omega^2(\mu_2 + \nu_2)),$$

assumption (3.26) is satisfied. Therefore, one obtains the upper bound (3.27) for the risk with $A = s$ if $f \in \Xi_s(B)$ and $A = s + 1/2$ if $f \in \Omega_s(B)$. Here, $\Omega_s(B)$ and $\Xi_s(B)$ are defined by (2.1) and (3.22), respectively. Therefore, the upper bounds for the risk as well as the optimal values for bandwidths are provided by formula (2.11) with $d = b = 0$, $A_1 = A$ and $A_2 = \alpha - 1$. The proof that the lower bounds are given by formula (2.20) with $d = b = 0$ and $a = 1$ can be carried out similarly to the proof of Theorem 3 in Section 2.3 and Theorem 5 below. One can check that convergence rates coincide with the ones derived in Dattner *et al.* (2011). Adaptive choice of the bandwidth by the Lepskii method is described in details in Dattner *et al.* (2011).

Example 3 (continuation). Estimation of generalized moments. Consider estimation of functional of the form (3.9) where $u(\theta) = (\theta^2 + 1)^{-1}$ and $m \geq 2$. It is easy to see that, although $u(\theta)$ is absolutely and square integrable, function $\theta^m (\theta^2 + 1)^{-1}$ is not. Then, $u^*(\omega) = \pi e^{-|\omega|}$ and, by formula (3.9), obtain

$$\Phi_u = \int_{-\infty}^{\infty} \theta^m (\theta^2 + 1)^{-1} f(\theta) d\theta = i^{-m} \int_{-\infty}^{\infty} e^{-|\omega|} \frac{d^m f^*(\omega)}{d\omega^m} d\omega. \quad (3.28)$$

Hence, formula (3.28) is a particular type of (3.10) with $\psi_{m1}^*(\omega) = 2e^{-\omega}$ and $\psi_{m2}^* = 0$ if m is even and $\psi_{m1}^*(\omega) = 0$ and $\psi_{m2}^* = 2e^{-\omega}$ if m is odd.

Note also that, since function $u^*(\omega)$ does not have a singularity at zero, construction of the estimator does not require partition of Φ_u into the parts corresponding to $|\omega| \leq 1$ and $|\omega| > 1$. Instead, applying integration by parts directly to the right-hand side of (3.28), obtain

$$\Phi_u = i^{-m} \left[\int_{-\infty}^{\infty} e^{-|\omega|} [\text{sign}(\omega)]^m f^*(\omega) d\omega - 2 \sum_{j=0}^{\frac{m-2}{2}} (f^*)^{(m-2-2j)}(0) \right]. \quad (3.29)$$

Using the general Leibniz formula for the l -th derivative of the product, write $(f^*)^{(l)}(0)$ as

$$(f^*)^{(l)}(0) = \sum_{k=0}^l \binom{l}{k} i^k \mu_k \frac{d^{l-k}}{d\omega^{l-k}} \left[\frac{1}{g^*(\omega)} \right]_{\omega=0}$$

where μ_k is defined in (3.3). Subsequently, estimate Φ_u by $\widehat{\Phi}_{u,h} = i^{-m}(\widehat{\Phi}_{u,1,h} - 2\widehat{\Phi}_{u,2})$ where

$$\begin{aligned}\widehat{\Phi}_{u,1,h} &= \int_{-1/h}^{1/h} e^{-|\omega|} [\text{sign}(\omega)]^m \frac{\widehat{q}^*(\omega)}{g^*(\omega)} d\omega, \\ \widehat{\Phi}_{u,2} &= \sum_{j=0}^{\frac{m-2}{2}} \sum_{k=0}^{m-2-2j} \binom{m-2-2j}{k} i^k \widehat{\mu}_k \frac{d^{m-2-2j-k}}{d\omega^{m-2-2j-k}} \left[\frac{1}{g^*(\omega)} \right]_{\omega=0}\end{aligned}$$

$\widehat{\mu}_k = \sum_{l=1}^n Y_l^k$ and $\widehat{q}^*(\omega)$ is defined in (2.8).

Let f and g satisfy conditions (3.23). Let also g^* satisfy Assumption A1 and inequality (3.24). Then $\sigma_{jk}^2(\omega)$ are uniformly bounded for $j = 0, \dots, m$, $k = 1, 2$, and inequality (3.26) holds. Moreover, (3.25) is valid with $a_m = 0$ and $b = d = 1$. Then, Theorem 4 yields the upper bounds for the risk of the form (3.27) where $A = s$ if $f \in \Xi_s(B)$ and $A = s + 1/2$ if $f \in \Omega_s(B)$. Applying formula (2.11) of Lemma 1 with modifications carried out in Theorem 2, obtain the following upper bounds for the minimax risk $\widehat{R}_n \equiv R_n(\widehat{\Phi}_{u,\widehat{h}_n}, \Omega_s(B))$ of the adaptive estimator $\widehat{\Phi}_{u,\widehat{h}_n}$:

$$\widehat{R}_n \leq \begin{cases} Cn^{-1}, \widehat{h}_n = 0 & \text{if } \beta < 1 \text{ or } \beta = 1, \gamma < 1, \\ Cn^{-1} (\log n)^{2\alpha+1}, \widehat{h}_n = (\log n)^{-1}, & \text{if } \beta = 1, \gamma = 1, \\ Cn^{-\frac{1}{\gamma}} (\log n)^{-V_1}, \widehat{h}_n = [\log n / (2\gamma)]^{-1}, & \text{if } \beta = 1, \gamma > 1, \\ C(\log n)^{-2A} \exp \left\{ -2 \left(\frac{\log n}{2\gamma} \right)^{\frac{1}{\beta}} \right\}, \widehat{h}_n = [\log n / (2\gamma)]^{-1}, & \text{if } \beta > 1, \gamma > 0. \end{cases}$$

The lower bounds for the minimax risk can be obtained similarly to Theorem 3 and coincide with the minimax upper bounds when $\beta < 1$ or $\beta = 1, \gamma < 1$, and differ from it by a logarithmic factor of n otherwise.

Example 4. Estimation of the $(2M + 1)$ -th absolute moment of the deconvolution density. Consider estimation of a functional of the form

$$\Phi_{2M+1} = \int_{-\infty}^{\infty} |\theta|^{2M+1} f(\theta) d\theta. \quad (3.30)$$

Since $|\theta|^{2M+1} = \theta^{2M+1} \text{sign}(\theta)$, functional (3.30) is a particular case of $\Phi(m, \theta_0)$ given by (3.7) with $m = 2M + 1$ and $\theta_0 = 0$:

$$\Phi_{2M+1} = (-1)^{M+1} \frac{2}{\pi} \int_0^{\infty} \frac{1}{\omega} \frac{d^{2M+1}}{d\omega^{2M+1}} \Re[f^*(\omega)] d\omega,$$

i.e., Φ_{2M+1} is of the form (3.10) with $\psi_{2M+1,1}^*(\omega) = 2(-1)^{M+1}(\pi\omega)^{-1}$ and $\psi_{2M+1,2}^*(\omega) = 0$. Similarly to (3.11), re-write Φ_{2M+1} as $\Phi_{2M+1} = \Phi_{2M+1,1} + \Phi_{2M+1,2}$ where $\Phi_{2M+1,1}$ and $\Phi_{2M+1,2}$ are the portions of Φ_{2M+1} evaluated over intervals $[0, 1]$ and $(1, \infty)$. Here,

$$\Phi_{2M+1,1} = (-1)^{M+1} \frac{2}{\pi} \sum_{j=0}^{2M+1} \binom{2M+1}{j} \int_0^1 \frac{1}{\omega} \frac{d^{2M+1-j}}{d\omega^{2M+1-j}} \left[\frac{1}{g^*(\omega)} \right] u_{j1}(\omega) d\omega,$$

where $u_{j1}(\omega)$ are defined in (3.13). Taking into account relations (3.15) and partitioning the sum above into the portions with the even and the odd indices, we obtain the following estimator of

$\Phi_{2M+1,1}$

$$\begin{aligned}\widehat{\Phi}_{2M+1,1} &= \frac{2}{\pi} \sum_{k=0}^M (-1)^{M+k} \left\{ \binom{2M+1}{2k+1} \int_0^1 \frac{\widehat{v}_{2k+1,2}(\omega)}{\omega} \frac{d^{2(M-k)}}{d\omega^{2(M-k)}} \left[\frac{1}{g^*(\omega)} \right] d\omega \right. \\ &\quad \left. - \binom{2M+1}{2k} \int_0^1 \frac{\widehat{v}_{2k,1}(\omega)}{\omega} \frac{d^{2(M-k)+1}}{d\omega^{2(M-k)+1}} \left[\frac{1}{g^*(\omega)} \right] d\omega \right\}.\end{aligned}\quad (3.31)$$

Note that for $\sigma_{j1}^2(\omega)$ defined in (3.21), one has

$$\sigma_{2j,1}^2(\omega) \asymp \text{Var}[Y^{2j} \cos(\omega Y)] \leq \int_{-\infty}^{\infty} y^{4j} q(y) dy, \quad (3.32)$$

$$\sigma_{2j+1,2}^2(\omega) \asymp \text{Var}[Y^{2j+1} \sin(\omega Y)] \leq \min \left[\int_{-\infty}^{\infty} y^{4j+2} q(y) dy, \omega^2 \int_{-\infty}^{\infty} y^{4j+4} q(y) dy \right]. \quad (3.33)$$

Hence, condition (3.26) is guaranteed by $\mu_{4M+4} < \infty$ and $\nu_{4M+4} < \infty$ where μ_k and ν_k are defined in (3.3).

Consider now $\widehat{\Phi}_{2M+1,2}$. Taking into account that, for any $j = 0, 1, 2, \dots$, one has

$$\frac{d^j \psi_{2M+1,1}^*(\omega)}{d\omega^j} = \frac{2}{\pi} \frac{(-1)^{M+1+j} j!}{\omega^{j+1}},$$

apply (3.19) and (3.20) for this particular case to obtain

$$\begin{aligned}\widehat{\Phi}_{2M+1,2,h} &= \frac{2}{\pi} \left\{ (-1)^M \sum_{k=1}^{2M+1} \sum_{j=0}^{2M+1-k} (k-1)! \binom{2M+1-k}{j} \frac{d^{2M+1-k-j}}{d\omega^{2M+1-k-j}} \left[\frac{1}{g^*(\omega)} \right]_{\omega=1} \widehat{u}_{j1}(1) \right. \\ &\quad \left. - \int_1^{1/h} \widehat{v}_{01}(\omega) \omega^{-(2M+2)} [g^*(\omega)]^{-1} d\omega \right\}\end{aligned}\quad (3.34)$$

Finally, Φ_{2M+1} can be estimated as $\widehat{\Phi}_{2M+1,h} = \widehat{\Phi}_{2M+1,1} + \widehat{\Phi}_{2M+1,2,h}$ where $\widehat{\Phi}_{2M+1,1}$ and $\widehat{\Phi}_{2M+1,2,h}$ are given by (3.31) and (3.34), respectively. Note that condition (3.25) holds with $m = 2M+1$, $a_m = 1$ and $b = d = 0$.

In particular, if $M = 0$, formulae (3.31) and (3.34) yield the following estimator for the first absolute moment of f

$$\widehat{\Phi}_{1,1} = \frac{2}{\pi} \int_0^1 \left(\frac{\widehat{v}_{01}(\omega) (g^*)'(\omega)}{\omega (g^*(\omega))^2} + \frac{\widehat{v}_{12}(\omega)}{\omega g^*(\omega)} \right) d\omega, \quad (3.35)$$

$$\widehat{\Phi}_{1,2,h} = \frac{2}{\pi} \frac{\widehat{v}_{01}(1)}{g^*(1)} - \frac{2}{\pi} \int_1^{1/h} \frac{1}{\omega^2} \frac{\widehat{v}_{01}(\omega)}{g^*(\omega)} d\omega. \quad (3.36)$$

and $\widehat{\Phi}_{1,h} = \widehat{\Phi}_{1,1} + \widehat{\Phi}_{1,2,h}$.

In order to derive upper and lower bounds for the minimax risk of the estimator $\widehat{\Phi}_{2M+1,h}$, we introduce the following sets of functions:

$$\Xi_s(B_1, B_2) = \left\{ f : \int_{-\infty}^{\infty} \theta^{4M+4} f(\theta) d\theta \leq B_1, \quad \sup_{\omega} [|f^*(\omega)| (|\omega|^s + 1)] \leq B_2 \right\} \quad (3.37)$$

Theorem below provides upper and lower bounds for the minimax risk of an estimator of Φ_{2M+1} in (3.30) while Corollary 3 produces similar results for the discrete version of the functional $\Phi_{2M+1,n} = n^{-1} \sum_{i=1}^n |\theta_i|^{2M+1}$.

Theorem 5 Let $\mu_{4M+4} \leq B_g < \infty$ for some positive constant B_g where μ_k is defined in (3.3). Let function g^* satisfy condition (3.24) and also be such that

$$\sup_{|\omega| \leq 1} \left| \frac{1}{\omega} \frac{d^{2k+1}}{d\omega^{2k+1}} \left(\frac{1}{g^*(\omega)} \right) \right| \leq C_{gM}, \quad 0 \leq k \leq M. \quad (3.38)$$

If inequality (2.2) in Assumptions A1 holds, then

$$R_n(\widehat{\Phi}_{2M+1, \tilde{h}_n}, \Xi_s(B_1, B_2)) \leq \begin{cases} C n^{-1}, \tilde{h}_n = 0 & \text{if } 2\alpha - 4M < 3, \beta = \gamma = 0 \\ C n^{-1} \log n, \tilde{h}_n = n^{-\frac{1}{4M+2}} & \text{if } 2\alpha - 4M = 3, \beta = \gamma = 0 \\ C n^{-\frac{4M+2s+2}{2s+2\alpha-1}}, \tilde{h}_n = n^{-\frac{1}{2s+2\alpha-1}}, & \text{if } 2\alpha - 4M > 3, \beta = \gamma = 0 \\ C (\log n)^{-\frac{4M+2s+2}{\beta}}, \hat{h}_n = \tilde{h}_n^{**} & \text{if } \beta > 0, \gamma > 0, \end{cases} \quad (3.39)$$

where, $\tilde{h}_n^{**} = \left[\frac{1}{2\gamma} \left(\log n - \frac{2s+2\alpha-1-\beta}{\beta} \log \log n \right) \right]^{-\frac{1}{\beta}}$.

If inequality (2.3) in Assumptions A1 holds, then

$$R_n(\Xi_s(B_1, B_2)) \geq \begin{cases} C n^{-1} & \text{if } 2\alpha - 4M \leq 3, \beta = \gamma = 0 \\ C n^{-\frac{4M+2s+2}{2s+2\alpha-1}} & \text{if } 2\alpha - 4M > 3, \beta = \gamma = 0 \\ C (\log n)^{-\frac{4M+2s+2}{\beta}} & \text{if } \beta > 0, \gamma > 0 \end{cases} \quad (3.40)$$

Here, $R_n(\widehat{\Phi}_{2M+1, \tilde{h}_n}, \Xi_s(B_1, B_2)) = \sup_{f \in \Xi_s(B_1, B_2)} \mathbb{E}(\widehat{\Phi}_{2M+1, \tilde{h}_n} - \Phi_{2M+1})^2$ and $R_n(\Xi_s(B_1, B_2)) = \inf_{\tilde{\Phi}} \sup_{f \in \Omega_s(B_1, B_2)} \mathbb{E}(\tilde{\Phi} - \Phi_{2M+1})^2$.

Note that the values of \tilde{h}_n are independent of unknown parameters s, B_1 and B_2 if $2\alpha - 4M \leq 3$ and $\beta = \gamma = 0$. If $\beta > 0, \gamma > 0$, one can replace \tilde{h}_n by $\hat{h}_n = [\log n / (3\gamma)]^{-1/\beta}$ and the rates of convergence will not change. Finally, if $2\alpha - 4M > 3$ and $\beta = \gamma = 0$, one can find the value of \hat{h}_n using the Lepskii method similarly to the regular case with the price of $\log n$ in the convergence rates.

Corollary 3 Let $\theta_i, i = 1, \dots, n$, in (1.2) be i.i.d. with pdf f . If $\Phi_{2M+1, n}$ is defined by formula (1.3) with $\varphi(\theta) = |\theta|^{2M+1}$, then under assumptions of Theorem 5, for sufficiently large n , one has

$$R_n(\widehat{\Phi}_{2M+1, \tilde{h}_n}, \Phi_{2M+1, n}, \Xi_s(B_1, B_2)) = \sup_{f \in \Xi_s(B)} \mathbb{E}(\widehat{\Phi}_{2M+1, \tilde{h}_n} - \Phi_{2M+1, n})^2 \asymp R_n(\widehat{\Phi}_{2M+1, \tilde{h}_n}, \Xi_s(B_1, B_2)),$$

$$R_n(\Phi_{2M+1, n}, \Xi_s(B_1, B_2)) = \inf_{\tilde{\Phi}_n} \sup_{f \in \Xi_s(B)} \mathbb{E}(\tilde{\Phi}_n - \Phi_{2M+1, n})^2 \asymp R_n(\Xi_s(B_1, B_2)),$$

where $R_n(\widehat{\Phi}_{2M+1, \tilde{h}_n}, \Xi_s(B_1, B_2))$ and $R_n(\Xi_s(B_1, B_2))$ are given by (3.39) and (3.40), respectively.

Remark 2 The choice of the class of functions. Observe that we derived the lower and the upper bounds for the risk not for the subset of the Sobolev ball $\Omega_s(B_2)$ but rather for the subset of $\Xi_s(B_2)$. This is motivated by our intention to compare our estimator for the first absolute moment with the respective estimator of Cai and Low (2011). One can easily obtain upper and lower bounds for the minimax risk over the set $\Omega_s(B_1, B_2)$ in a very similar manner.

Remark 3 Relation to Cai and Low (2011). Cai and Low (2011) studied estimation of Φ_n of the form (1.3) with $\varphi(\theta) = |\theta|$ based on data generated by model (1.2) where the errors ξ_i are i.i.d. $\mathcal{N}(0, \sigma^2)$ and there are no probabilistic assumptions on vector θ . They showed that

$$\inf_{\tilde{\Phi}} \sup_{\theta \in \Theta_n(M_0)} \mathbb{E}(\tilde{\Phi}_n - \Phi)^2 \asymp M_0^2 \left(\frac{\log \log n}{\log n} \right)^2, \quad (3.41)$$

$$\inf_{\tilde{\Phi}} \sup_{\theta \in \mathbb{R}^n} \mathbb{E}(\tilde{\Phi}_n - \Phi)^2 \asymp \frac{1}{\log n},$$

where $\Theta_n(M_0) = \{\theta : \|\theta\|_\infty \leq M_0\}$. By employing a state of the art procedure based on Chebyshev and Hermite polynomials, they constructed adaptive estimators that attain these convergence rates. With the assumption that θ_i are generated independently from pdf f , the problem reduces to estimation of Φ_1 in (3.30), the first absolute moment of the mixing density. Using formulae (3.35) and (3.36), one can construct an estimator $\hat{\Phi}_1$ of Φ_1 .

Note that, since in the case of Gaussian errors, one has $\alpha = 0$, $\beta = 2$ and $\gamma = \sigma^2/2$, the estimators (3.35) and (3.36) are adaptive if $h = \hat{h}_n = [2 \log n / (3\sigma^2)]^{-1/2}$, and Corollary 3 implies that

$$R_n(\hat{\Phi}_{\hat{h}_n}, \Phi_n, \Xi_s(B_1, B_2)) \asymp R_n(\Phi_n, \Xi_s(B_1, B_2)) \asymp (\log n)^{-(s+1)}$$

where $\Xi_s(B_1, B_2)$ is defined in (3.37). Since f is a pdf, it is absolutely integrable, so that $s \geq 0$. Therefore, convergence rate (3.41) corresponds to “the worst case scenario” where $s = 0$ and f is a combination of delta functions. The estimator of Cai and Low (2011) addresses this “worst-case scenario” but is unable to adapt to a more favorable situation where $|f^*(\omega)| \rightarrow 0$ as $|\omega| \rightarrow \infty$. In addition, our estimator is more flexible since it is constructed for any type of error density g .

On the other hand, the estimator of Cai and Low (2011) does not impose any probabilistic assumptions on θ_i , and, therefore, can be advantageous when the values of θ_i , $i = 1, \dots, n$, are not independent.

4 The sparse case

4.1 Estimation procedure and the upper bounds for the risk

The objective of this section is to estimate the functional $\Phi_\mu = \int_{-\infty}^{\infty} \varphi(x) f_0(x) dx$ defined by (1.5) where $f_0(\theta)$ is pdf of the nonzero entries of θ and

$$f(x) = \mu_n f_0(x) + (1 - \mu_n) \delta(x), \quad (4.1)$$

where μ_n is known. Discrete version of (4.1) appears as the problem of estimation of the functional Φ_{k_n} defined by formula (1.4), where the average number $k_n = n^\nu$, $0 < \nu < 1$, of nonzero entries of vector θ in (1.2) is known but locations of the zero entries of θ are not. Observe that $k_n = n^\nu$ corresponds to $\mu_n = n^{-1} k_n = n^{\nu-1}$ in (4.1). Again, similarly to the non-sparse case, as long as $\mathbb{E}|\varphi(\theta)|^2 < \infty$, one has $\mathbb{E}(\Phi_{k_n} - \Phi_\mu)^2 \leq k_n^{-1} \mathbb{E}|\varphi(\theta)|^2$, so that the minimax errors for estimating Φ_{k_n} and Φ_μ are equivalent up to the $C k_n^{-1}$ additive term.

Due to (4.1), one has $\Phi = \mu_n \Phi_\mu + (1 - \mu_n) \varphi(0)$, so that the value of Φ_μ can be recovered as

$$\Phi_\mu = \mu_n^{-1} \Phi - \mu_n^{-1} (1 - \mu_n) \varphi(0). \quad (4.2)$$

Therefore, we estimate Φ_μ by

$$\hat{\Phi}_{\mu, h} = \frac{\hat{\Phi}_h}{\mu_n} - \frac{1 - \mu_n}{\mu_n} [\varphi(0) - \delta_h] \quad \text{with} \quad \delta_h = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi^*(-\omega) \mathbb{I}(|\omega| > h^{-1}) d\omega, \quad (4.3)$$

where $\widehat{\Phi}_h$ is defined in (2.7) and the correction term δ_h is a completely known non-random quantity.

In order to justify estimator (4.3), we derive expressions for its variance and bias. Since the second term in (4.3) is non-random, $\text{Var}(\widehat{\Phi}_{\mu,h}) = \mu_n^{-2} \text{Var}(\widehat{\Phi}_h)$ where $\text{Var}(\widehat{\Phi}_h)$ is bounded by

$$\text{Var}(\widehat{\Phi}_{\mu,h}) \leq \frac{\|g\|_\infty}{2\pi n \mu_n^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2} \mathbb{I}(|\omega| \leq h^{-1}) d\omega.$$

Since $f^*(\omega) = \mu_n f_0^*(\omega) + (1 - \mu_n)$, the bias term of $\widehat{\Phi}_{\mu,h}$ is of the form

$$\begin{aligned} \mathbb{E}\widehat{\Phi}_{\mu,h} - \Phi_\mu &= \frac{1}{2\pi\mu_n} \int_{-\infty}^{\infty} \varphi^*(-\omega) f^*(\omega) \mathbb{I}(|\omega| \leq h^{-1}) d\omega \\ &- \frac{1 - \mu_n}{\mu_n} [\varphi(0) - \delta_h] - \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi^*(-\omega) f_0^*(\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi^*(-\omega) f_0^*(\omega) \mathbb{I}(|\omega| > h^{-1}) d\omega + \frac{1 - \mu_n}{2\pi\mu_n} \Delta(h) \end{aligned}$$

where

$$\Delta(h) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi^*(-\omega) \mathbb{I}(|\omega| \leq h^{-1}) d\omega - \varphi(0) + \delta_h = 0.$$

Therefore, for any $f_0 \in \Omega_s(B)$, where $\Omega_s(B)$ is defined in (2.1), one derives:

$$(\mathbb{E}\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 \leq \frac{B^2}{4\pi^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2 + 1)^s} \mathbb{I}(|\omega| > h^{-1}) d\omega.$$

Hence,

$$\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 \leq \frac{\|g\|_\infty}{2\pi n \mu_n^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{|g^*(\omega)|^2} \mathbb{I}(|\omega| \leq h^{-1}) d\omega + \frac{B^2}{4\pi^2} \int_{-\infty}^{\infty} \frac{|\varphi^*(\omega)|^2}{(\omega^2 + 1)^s} \mathbb{I}(|\omega| > h^{-1}) d\omega. \quad (4.4)$$

Let $n_\mu = n\mu_n^2 = n^{-1}k_n^2$ be the new, "effective" sample size. Then, comparing (4.4) with (2.9), one immediately observes that the upper bounds for the risk of the estimator $\widehat{\Phi}_{\mu,h}$ of Φ_μ would coincide with the upper bounds for the risk of the estimator $\widehat{\Phi}_h$ of Φ in the non-sparse case if the sample size n were replaced by the effective sample size n_μ . Denote

$$R_{n,\mu_n}(\widehat{\Phi}_{\mu,h}, \Omega_s(B)) = \sup_{f_0 \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 \quad (4.5)$$

where $\Omega_s(B)$ is defined in (2.1). If $\nu > 1/2$, then $n_\mu = n\mu_n^2 = n^{2\nu-1} > 1$, so that combination of Theorem 1 and formula (4.4) immediately yields the upper bounds for the risk.

Theorem 6 *Let g be bounded above and observations be given by model (1.2) where f is of the form (4.1) with $\mu_n = n^{1-\nu}$, $\nu > 1/2$. Then, under Assumptions A1 and A2 (inequalities (2.2) and (2.5) only), one derives*

$$R_{n,\mu_n}(\widehat{\Phi}_{\mu,\check{h}_n}, \Omega_s(B)) \leq C \Delta_{n^{1-2\nu}}(s + a + (b - 1)/2, \alpha - a, b, d, \beta, \gamma) \quad (4.6)$$

where the expressions for $\check{h}_n = \check{h}_{n^{2\nu-1}}$ and $\Delta_{n^{2\nu-1}}(A_1, A_2, b, d, \beta, \gamma)$ are given by (2.11), with n replaced by $n^{2\nu-1}$. In addition, one has the following adaptive minimax convergence rates $\widehat{R_{n,\mu_n}} \equiv$

$$R_n(\widehat{\Phi}_{\widehat{h}_n}, \Omega_s(B))$$

$$\begin{aligned}
\widehat{R_{n,\mu_n}} &\asymp n^{-(2\nu-1)}, \widehat{h}_n = 0 & \text{if } b > \beta, \\
\widehat{R_{n,\mu_n}} &\asymp n^{-(2\nu-1)}, \widehat{h}_n = 0 & \text{if } b = \beta, d > \gamma > 0 \\
\widehat{R_{n,\mu_n}} &\asymp n^{-(2\nu-1)}, \widehat{h}_n = 0 & \text{if } b = \beta, d = \gamma, a > \alpha + 1/2 \\
\widehat{R_{n,\mu_n}} &\asymp n^{-(2\nu-1)} \log \log n, \widehat{h}_n = \widehat{h}_n^* & \text{if } b = \beta > 0, d = \gamma > 0, a = \alpha + 1/2 \\
\widehat{R_{n,\mu_n}} &\asymp n^{-(2\nu-1)} \log n, \widehat{h}_n = n^{-\frac{1}{2a-1}} & \text{if } b = \beta = 0, d = \gamma = 0, a = \alpha + 1/2 \\
\widehat{R_{n,\mu_n}} &\asymp n^{-(2\nu-1)} (\log n)^{\frac{2\alpha-2a+1}{\beta}}, \widehat{h}_n = \widehat{h}_n^* & \text{if } b = \beta > 0, d = \gamma > 0, a < \alpha + 1/2 \\
\widehat{R_{n,\mu_n}} &\asymp n^{-\frac{(2\nu-1)(2s+2a-1)}{2s+2\alpha}} \log n, \widehat{h}_n = h_{\widehat{j}} & \text{if } b = \beta = 0, d = \gamma = 0, a < \alpha + 1/2 \\
\widehat{R_{n,\mu_n}} &\asymp (\log n)^{-\frac{U_1}{\beta}} n^{-\frac{d(2\nu-1)}{\gamma}}, \widehat{h}_n = \widehat{h}_n^* & \text{if } b = \beta > 0, \gamma > d > 0 \\
\widehat{R_{n,\mu_n}} &\asymp (\log n)^{-\frac{U_2}{\beta}} \exp\left(-2d \left[\frac{\log n}{2\gamma}\right]^{b/\beta}\right), \widehat{h}_n = \widehat{h}_n^* & \text{if } \beta > b > 0, d > 0, \gamma > 0, \\
\widehat{R_{n,\mu_n}} &\asymp (\log n)^{-\frac{2s+2a-1}{\beta}}, \widehat{h}_n = \widehat{h}_n^{**} & \text{if } b = d = 0, \beta > 0, \gamma > 0
\end{aligned} \tag{4.7}$$

Here, $\widehat{h}_n^* = [(2\nu - 1) \log n / (2\gamma)]^{-\frac{1}{\beta}}$, $\widehat{h}_n^{**} = (\log n / 4\gamma)^{-\frac{1}{\beta}}$, \widehat{j} is defined in (2.15) with C_Φ given by (2.16) and n replaced by $n^{2\nu-1}$, $U_1 = \min(\beta + 2a + 2s - 1, \beta + 2a - 2\alpha - 1)$ and $U_2 = b + 2a + 2s - 1$.

4.2 The lower bounds for the risk

The upper bounds for the risk in formula (4.4) and Theorem 6 suggest that, for $0 \leq \nu \leq 1/2$, one has $n_\mu = n\mu_n^2 = n^{2\nu-1} \leq 1$ and construction of a consistent estimator is impossible for any functional of the form (1.5). The next proposition shows that this, indeed, is true in a wide variety of situations. In particular, under mild assumptions, the risk of no estimator can converge to zero faster than $C(n\mu_n^2)^{-1}$.

Theorem 7 *Let $f(x)$ be given by (4.1) and Φ_μ be defined by (1.5). If, for some $C_I > 0$, there exist two pdfs, $f_1(\theta)$ and $f_2(\theta)$ such that*

$$I_k = \int_{-\infty}^{\infty} g^{-1}(x) \left[\int_{-\infty}^{\infty} g(x - \theta) f_k(\theta) d\theta \right]^2 dx \leq C_I < \infty, \quad k = 1, 2, \tag{4.8}$$

and

$$\Delta_{12} = \int_{-\infty}^{\infty} \varphi(\theta) [f_1(\theta) - f_2(\theta)] d\theta \neq 0, \tag{4.9}$$

then

$$\inf_{\widetilde{\Phi}_\mu} \sup_{f_1, f_2} \mathbb{E}(\widetilde{\Phi}_\mu - \Phi_\mu)^2 \geq C \min(1, (n\mu_n^2)^{-1}), \tag{4.10}$$

where $\widetilde{\Phi}_\mu$ is any estimator of Φ_μ based on observations Y_1, \dots, Y_n .

Theorem 7 implies that, although the proof of the low bounds for the risk in Cai and Low's (2004, 2011) depend heavily on the normality assumption, the fact that one needs $\nu > 1/2$ in order to consistently estimate Φ_μ remains valid, whether ξ_i in (1.2) are normally distributed or not. Moreover, Theorem 7 does not require function φ to be integrable or square integrable, so one can apply this theorem easily to a wide variety of functionals. The quantity $n_\mu^{-1} = (n\mu_n^2)^{-1}$ acts as the parametric convergence rate that cannot be surpassed. Corollary 4 below shows that Theorem 7 holds in the case when g is a Gaussian pdf. A similar calculation can be repeated in the case when, for example, g is a doubly-exponential pdf.

Corollary 4 Let $f(x)$ be given by (4.1) and Φ be defined by (1.5). If $g(x) = \mathcal{N}(x|0, \sigma^2)$ is a Gaussian pdf and $\int_{-\infty}^{\infty} |\varphi(x)|g(x)dx < \infty$, then the lower bound (4.10) holds.

Corollary 4 generalizes the results of Cai and Low who proved the lower bound (4.10) in the case of Gaussian errors if $\varphi(x) = x$ (Cai and Low (2004)) and $\varphi(x) = |x|$ (Cai and Low (2011)).

One would like to derive the lower bounds for the risk of the form (2.20) for any combination of function $\varphi(\theta)$ and $g(\theta)$. Unfortunately, we succeeded in doing this only in the cases when $g(\theta)$ has polynomial descent as $|\theta| \rightarrow \infty$ (Theorem 8) or when $d = 0$ and $\gamma > 0$ in Assumptions A1 and A2 (Theorem 9). Denote

$$R_{n, \mu_n}(\Omega_s(B)) = \inf_{\tilde{\Phi}_\mu} \sup_{f_0 \in \Omega_s(B)} \mathbb{E}(\tilde{\Phi}_\mu - \Phi_\mu)^2 \quad (4.11)$$

where $\tilde{\Phi}_\mu$ is any estimator of Φ_μ based on observations Y_1, \dots, Y_n and $\Omega_s(B)$ is defined in (2.1). Then, the following theorem is true.

Theorem 8 Let $f(x)$ be given by (4.1) and Φ be defined by (1.5). Let g be bounded above and such that $|g(\theta)| \geq C_{g1}(\theta^2 + 1)^{-\varsigma}$. Let function g^* be ς_0 times continuously differentiable, where ς_0 is the closest integer no less than ς , and satisfy the following condition

$$\frac{|d^l g^*(\omega)|}{d\omega^l} \leq C_{g2} |g^*(\omega)| (1 + |\omega|)^{\tau_l}, \quad l = 1, 2, \dots, \varsigma_0, \quad \text{where } \tau = 0 \quad \text{if } \gamma = 0. \quad (4.12)$$

Let there exist $\omega_0 \in (0, \infty)$ such that function $\rho(\omega) = \arg(\varphi^*(\omega))$ is ς_0 times continuously differentiable for $|\omega| \geq \omega_0$, with $|\rho^{(j)}(\omega)| \leq \rho < \infty$, $j = 0, 1, 2, \dots, \varsigma_0$. Then, under Assumptions A1 and A2 (inequalities (2.3) and (2.4) only), when $\mu_n = n^{\nu-1}$ with $\nu > 1/2$, one derives

$$R_{n, \mu_n}(\Omega_s(B)) \geq \begin{cases} C n^{-(2\nu-1)} & \text{if } b > \beta, \\ C n^{-(2\nu-1)} & \text{if } b = \beta, d > \gamma > 0 \\ C n^{-(2\nu-1)} & \text{if } b = \beta, d = \gamma, a \geq \alpha + 1/2 \\ C n^{-\frac{(2\nu-1)(2s+2a-1)}{2s+2\alpha}} & \text{if } d = \gamma = 0, a < \alpha + 1/2 \\ C (\log n)^{-\frac{U_5}{\beta}} n^{-\frac{d(2\nu-1)}{\gamma}} & \text{if } b = \beta, \gamma > d > 0 \\ C (\log n)^{-\frac{U_6}{\beta}} \exp\left(-2d \left[\frac{\log n(2\nu-1)}{2\gamma}\right]^{b/\beta}\right) & \text{if } b < \beta, d > 0, \gamma > 0 \\ C (\log n)^{-\frac{2s+2a-1}{\beta}} & \text{if } b = d = 0, \beta > 0, \gamma > 0 \end{cases} \quad (4.13)$$

where $U_5 = 2a + 2s(1 - d/\gamma) - 2d\alpha/\gamma + 4\beta(\varsigma_0 + 1) - (U_{\tau, \varsigma_0} + 1)d/\gamma - 1$, $U_6 = 4b(\varsigma_0 + 1) + 2a + 2s - 1$ and $U_{\tau, \varsigma_0} = \min(\beta(4\varsigma_0 + 3) - 2\tau\varsigma_0 - 1, \beta(2\varsigma_0 + 3) + 2\varsigma_0 - 1)$.

Theorem 8 confirms that whenever g has polynomial descent, convergence rates in Theorem 6 are indeed optimal within, at most, a logarithmic factor of the sample size. Another general result refers to the case when $d = 0$ and $\gamma > 0$ in Assumptions A1 and A2.

Theorem 9 Let $f(x)$ be given by (4.1) and Φ_μ be defined by (1.5). Let $\mu_n = n^{\nu-1}$ with $\nu > 1/2$ and inequalities (2.3) and (2.4) in Assumptions A1 and A2 hold with $d = 0$ and $\gamma > 0$. Then,

$$R_{n, \mu_n}(\Omega_s(B)) \geq C (\log n)^{-\frac{2s+2a-1}{\beta}}. \quad (4.14)$$

Remark 4 Estimation of functionals of the form (1.4). Note that, for $1/2 < \nu < 1$, one has $k_n^{-1} = n^{-\nu} < n^{1-2\nu} = (n\mu_n^2)^{-1}$. Therefore, if θ_i , $i = 1, \dots, n$, in (1.2) are i.i.d. with the pdf f of the form (4.1), for any $h \geq 0$, one has one has

$$\begin{aligned}\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_{k_n})^2 &\leq 2\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 + 2k_n^{-1}\|\varphi\|_\infty^2 \leq \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 + 2n^{-(1-\nu)}(n\mu_n^2)^{-1}\|\varphi\|_\infty^2, \\ \mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_{k_n})^2 &\geq 1/2\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 - 2k_n^{-1}\|\varphi\|_\infty^2 \geq 1/2\mathbb{E}(\widehat{\Phi}_{\mu,h} - \Phi_\mu)^2 - 2n^{-(1-\nu)}(n\mu_n^2)^{-1}\|\varphi\|_\infty^2.\end{aligned}$$

Hence, if n is large enough, the upper and the minimax lower bounds for the risks of the estimators of Φ_{k_n} and Φ_μ coincide up to a constant factor. In particular, the following statement holds.

Corollary 5 *Let θ_i , $i = 1, \dots, n$, in (1.2) be i.i.d. with pdf f defined in (4.1). If $\varphi(\theta)$ is uniformly bounded $|\varphi(\theta)| \leq \|\varphi\|_\infty < \infty$, then under assumptions of Theorem 6, one has*

$$R_{n,\mu_n}(\widehat{\Phi}_{\mu,\tilde{h}_n}, \Phi_{k_n}, \Omega_s(B)) = \sup_{f_0 \in \Omega_s(B)} \mathbb{E}(\widehat{\Phi}_{\mu,\tilde{h}_n} - \Phi_{k_n})^2 \leq CR_{n,\mu_n}(\widehat{\Phi}_{\mu,\tilde{h}_n}, \Omega_s(B)),$$

where $R_{n,\mu_n}(\widehat{\Phi}_\mu, \Omega_s(B))$, defined in (4.5), is given by expression (4.6) and $C > 0$ is independent of n . Also, under assumptions of Theorem 8 (or Theorem 9), for sufficiently large n , one has

$$R_{n,\mu_n}(\Phi_{k_n}, \Omega_s(B)) = \inf_{\widehat{\Phi}_{k_n}} \sup_{f \in \Omega_s(B)} \mathbb{E}(\tilde{\Phi}_{k_n} - \Phi_{k_n})^2 \geq CR_{n,\mu_n}(\Omega_s(B)),$$

where $R_{n,\mu_n}(\Omega_s(B))$, defined in (4.11), is given by expression (4.13) (or (4.14)) and $C > 0$ is independent of n .

5 Simulation study

In order to evaluate small sample properties of the estimators presented in the paper we carried out a limited simulation study. In particular, we compared our estimator $\widehat{\Phi}_{1,h} = \widehat{\Phi}_{1,1} + \widehat{\Phi}_{1,2,h}$ of the first absolute moment of the mixing density Φ_1 , where $\widehat{\Phi}_{1,1}$ and $\widehat{\Phi}_{1,2,h}$ are defined in (3.35) and (3.36), respectively, with the estimator $\widehat{\Phi}_{CL}$ of Cai and Low (2011) which is based on approximation of the absolute value by combination of Chebyshev and Hermite polynomials:

$$\widehat{\Phi}_{CL} = \sum_{k=1}^{K_*} G_{2k}^* M_0^{1-2k} B_{2k}. \quad (5.1)$$

Here, M_0 is such that $|\theta_i| \leq M_0$, $i = 1, \dots, n$, G_{2k}^* is the coefficient for θ^{2k} in the expansion of $|\theta|$ by Chebyshev polynomials and

$$B_{2k} = n^{-1} \sum_{i=1}^n H_{2k}(y_i),$$

where $H_{2k}(x)$ are Hermite polynomials (with respect to $\exp(-x^2/2)$) of degree $2k$.

We considered three choices for the mixing density f : f is Gaussian $\mathcal{N}(m_\theta, \sigma_\theta^2)$, f is uniform on the interval $[a, b]$ and f is a combination of delta functions

$$f(x) = \sum_{k=1}^K b_k \delta(x - a_k). \quad (5.2)$$

Note that, in the latter case, variable θ is discrete and does not have a pdf in a regular sense. In particular, we chose $m_\theta = 2$ and $\sigma_\theta = 4$ when f is Gaussian, $a = -2$ and $b = 5$ when f is uniform

and $K = 3$, $a_1 = -2$, $a_2 = 0$, $a_3 = 5$ and $b_1 = b_2 = b_3 = 1/3$ when f is a combination of delta functions.

For the purpose of comparison, we generated the vector $\boldsymbol{\theta}$ using probability density function (or probability mass function) f and added a vector of independent errors $\boldsymbol{\xi}$ to it to obtain the vector of observations \mathbf{Y} . Since procedure of Cai and Low (2011) is developed for Gaussian errors only, we chose g to be the standard normal pdf $\mathcal{N}(0, 1)$. We considered four values of sample size, $n = 500$, $n = 200$, $n = 100$ and $n = 50$. For each of the combinations of vectors $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$, we evaluated

$$\Phi_n = n^{-1} \sum_{i=1}^n |\theta_i| \quad (5.3)$$

and obtained estimators $\hat{\Phi}_h$ and $\hat{\Phi}_{CL}$ based on observations \mathbf{Y} . While constructing estimator $\hat{\Phi}_h$ given by (2.7), we use the value $h = \hat{h}_n = [2 \log n / (3\sigma^2)]^{-1/2}$ with $\sigma = 1$, so that estimator $\hat{\Phi}_{\hat{h}_n}$ is fully adaptive. Estimator $\hat{\Phi}_{CL}$ depends on two parameters, K_* and M_0 . Following Cai and Low (2011), we chose $K_* = 0.5 \log n / \log \log n$. As far as parameter M_0 is concerned, we initially chose M_0 on the basis of our knowledge of f . In particular, we used $M_0 = \max(|a_k|)$ when f is of the form (5.2), $M_0 = \max(|a|, |b|)$ when f is uniform, and $M_0 = M_{n,0} = |m_\theta| + \sigma_\theta \sqrt{2 \log n}$ if f is Gaussian $\mathcal{N}(m_\theta, \sigma_\theta^2)$. Thus, the estimator $\hat{\Phi}_{CL}$ with those choices of M is not adaptive. Since in practice, the upper bound for the absolute values $|\theta_i|$ is not known, we also studied the case when f is of the form (5.2) and as it is suggested in Cai and Low (2011), $M_{n,0}$ is set to $M_{n,0} = \sqrt{2 \log n}$.

We repeated the process $N = 1000$ times. The accuracies of estimators are determined by their average squared deviations Δ and Δ_n from, respectively, Φ defined in (3.30) and Φ_n defined in (5.3):

$$\Delta = N^{-1} \sum_{i=1}^N (\hat{\Phi}_{\hat{h}_n}(i) - \Phi)^2, \quad \Delta_n = N^{-1} \sum_{i=1}^N (\hat{\Phi}_{\hat{h}_n}(i) - \Phi_n(i))^2.$$

Here $\hat{\Phi}_{\hat{h}_n}(i)$ and $\Phi_n(i)$ are the values of the estimator $\hat{\Phi}_{\hat{h}_n}$ and Φ_n in the i -th realization of the simulations. The standard deviations of Δ and Δ_n are listed below in parentheses. As a benchmark for the magnitude of the error, we also report the average squared deviation $s(\Phi_n)$ of $\Phi_n(i)$ from its mean Φ :

$$s(\Phi_n) = N^{-1} \sum_{i=1}^N (\Phi_n(i) - \Phi)^2.$$

Results of simulations are summarized in Table 1.

Simulations confirm that, as long as one uses the exact value of M_0 in (5.1), estimators $\hat{\Phi}_{\hat{h}_n}$ and $\hat{\Phi}_{CL}$ have very similar precisions: the difference between the average errors of the two estimators is smaller than respective standard deviations or the average squared deviation of Φ_n from its mean Φ . As it was expected, estimator $\hat{\Phi}_{CL}$ performs better when θ is a discrete random variable while the set up where θ is a continuous random variable benefits $\hat{\Phi}_{\hat{h}_n}$. Also, though $\hat{\Phi}_{\hat{h}_n}$ is designed to estimate Φ while $\hat{\Phi}_{CL}$ is intended to estimate Φ_n , the error Δ_n turns out to be smaller for $\hat{\Phi}_{\hat{h}_n}$ while $\hat{\Phi}_{CL}$ is somewhat more accurate as an estimator of Φ . The latter confirms that the problems are equivalent up to a small additive error and that both problems should be treated in a similar fashion.

However, the advantage of the estimator $\hat{\Phi}_{\hat{h}_n}$ is that it is adaptive: indeed we fixed the value of parameter h for all three distributions of θ . On the other hand, estimator $\hat{\Phi}_{CL}$ is not adaptive since parameter $M_0 = M_{n,0}$ is chosen on the basis of the pdf f which is not known. Our simulations show

COMPARISON OF THE ACCURACY OF THE FOURIER TRANSFORM BASED ESTIMATOR $\hat{\Phi}_{\hat{h}_n}$
AND THE CAI AND LOW (2011) ESTIMATOR $\hat{\Phi}_{CL}$ OVER 1000 SIMULATION RUNS

$f = [\delta(x+2) + \delta(x) + \delta(x-5)]/3, \Phi = 2.33333, M_0 = 5$ (true value)								
	n=500		n=200		n=100		n=50	
	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$
Δ	0.02365 (0.02664)	0.01292 (0.01698)	0.03727 (0.05428)	0.02659 (0.03612)	0.05631 (0.07795)	0.05225 (0.07583)	0.09774 (0.14302)	0.12955 (0.18895)
Δ_n	0.01606 (0.01204)	0.00583 (0.00574)	0.01834 (0.01862)	0.00816 (0.01016)	0.02174 (0.02803)	0.01228 (0.01681)	0.02938 (0.03974)	0.06771 (0.08313)
$s(\Phi_n)$	0.00854		0.02027		0.04168		0.08186	

f is uniform on $[-2, 5], \Phi = 2.07142, M_0 = 5$ (true value)								
	n=500		n=200		n=100		n=50	
	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$
Δ	0.00693 (0.00978)	0.04567 (0.03314)	0.01566 (0.02331)	0.06185 (0.05857)	0.03020 (0.04428)	0.07095 (0.08070)	0.06006 (0.08486)	0.05382 (0.07184)
Δ_n	0.00266 (0.00383)	0.04260 (0.01804)	0.00582 (0.00786)	0.04772 (0.02877)	0.01072 (0.01533)	0.04927 (0.04262)	0.02147 (0.03066)	0.02901 (0.03884)
$s(\Phi_n)$	0.00431		0.00992		0.01887		0.03868	

f is Gaussian $\mathcal{N}(m_\theta, \sigma_\theta^2), m_\theta = 2, \sigma_\theta = 4, \Phi = 3.58237, M_{n,0} = m_\theta + \sigma_\theta\sqrt{2\log n}$ (true value)								
	n=500		n=200		n=100		n=50	
	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$
Δ	0.01566 (0.02252)	0.01746 (0.02324)	0.04382 (0.06122)	0.03117 (0.04318)	0.08756 (0.12636)	0.06384 (0.09740)	0.16282 (0.22401)	0.24934 (0.27030)
Δ_n	0.00237 (0.00338)	0.01030 (0.00939)	0.00539 (0.00735)	0.00769 (0.01047)	0.01104 (0.01516)	0.01124 (0.01652)	0.01942 (0.02673)	0.22841 (0.16487)
$s(\Phi_n)$	0.01376		0.03881		0.08029		0.14125	

$f = [\delta(x+2) + \delta(x) + \delta(x-5)]/3, \Phi = 2.33333, M_{n,0} = \sqrt{2\log n}$								
	n=500		n=200		n=100		n=50	
	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$	$\hat{\Phi}_{\hat{h}_n}$	$\hat{\Phi}_{CL}$
Δ	0.02368 (0.02699)	1.05701 (0.25291)	0.03770 (0.04929)	2.40929 (0.81377)	0.06359 (0.08691)	4.76195 (2.20037)	0.09686 (0.13408)	1.11470 (1.06931)
Δ_n	0.01590 (0.01149)	1.05668 (0.26219)	0.01804 (0.01812)	2.44297 (0.91473)	0.02399 (0.02890)	4.83772 (2.54537)	0.03042 (0.04237)	0.91974 (0.62472)
$s(\Phi_n)$	0.00872		0.01986		0.04504		0.08724	

Table 1: The average values (over 1000 runs) of the errors Δ and Δ_n (with the standard deviations of the errors listed in the parentheses), and the average squared deviation $s(\Phi_n)$ of Φ_n from its mean Φ .

that estimator $\hat{\Phi}_{CL}$ is very sensitive to the choice of M_0 . Indeed, when, instead of the true value M_0 , we used $M_0 = M_{n,0} = \sqrt{2 \log n}$ suggested in Cai and Low (2011), precision of the estimator deteriorated so much that $\hat{\Phi}_{CL}$ stopped being consistent. The latter demonstrated the advantage of the estimator $\hat{\Phi}_{\hat{h}_n}$ in comparison with $\hat{\Phi}_{CL}$. Hence, although the estimator $\hat{\Phi}_{CL}$ is adequate in asymptotic setting, it runs into serious problems for moderate values of n that are more likely to occur in practical situations.

6 Discussion

This paper significantly advances the theory of estimation of linear functionals of the unknown deconvolution density. In particular, in the case when the Fourier transform of φ exists, we derived the general minimax lower bounds for the risk that have not been obtained earlier. We also elaborate on the upper bounds for the risk derived earlier by Butucea and Comte (2009). Using these results, we immediately retrieve the upper and the minimax lower bounds for the risk of the pointwise estimator of the mixing density with classical and Berkson errors studied by Delaigle (2007). Direct comparison with Delaigle (2007) shows that our upper risk bounds are more precise due to more flexible assumptions; they are also supplemented by the minimax lower bounds that have not been derived previously. In addition, our estimator is adaptive.

Furthermore, we expand our theory in a novel way to incorporate estimation of functionals of the form (1.1) where function $\varphi(x)$ does not have a Fourier transform in a regular sense. The new methodology relies on application of inversion formulas. We show that by using such formulas, many functionals of interest can be expressed as linear functionals of the real and the imaginary parts of the Fourier transform f^* of f and their derivatives. We construct estimators for such functionals and derive the upper bounds for their risks. As a particular case of application of our methodology, we automatically recover the estimators of the mixing cumulative distribution function investigated by Dattner *et al.* (2011), as well as their minimax lower and upper bounds for the risk. Other examples include estimation of the $(2M + 1)$ -th absolute moment of deconvolution density and the functional of the form $\int_{-\infty}^{\infty} \theta^m (\theta^2 + 1)^{-1} f(\theta) d\theta$ with $m \geq 2$. As a special case, we obtain an estimator of the first absolute moment of the deconvolution density (the problem studied by Cai and Low (2011) in the case of the Gaussian error distribution), and retrieve the upper and lower bounds for the minimax risk of this estimator as a particular case of our general minimax bounds.

Finally, we use the same approach to deal with the situation where deconvolution density f is delta-contaminated and is defined by (4.1). This set-up corresponds to the situation where the vector of unobservable variables is sparse and has, on the average, only k_n nonzero entries. Following Cai and Low (2011), we assume that k_n (or μ_n in (4.1)) is known and the objective is estimating the functional over these non-zero entries only. We show that convergence rates are determined by the “effective” sample size $n_\mu = n^{-1} k_n^2$, propose an estimation procedure specifically designed for the sparse case, and construct the minimax lower and the upper bounds for the risk. In particular, an immediate consequence of our theory is a proof that consistent estimation of any linear functional of deconvolution density is impossible if $k_n = n^\nu$ and $0 < \nu \leq 1/2$, since, in this case, the effective sample size $n_\mu = n^{2\nu-1} < 1$. This is a significant generalization of the results of Cai and Low (2011) who drew the same conclusion specifically for estimation of the first absolute moment of the unobservable variable under Gaussian errors.

We also carry out a limited simulation study which compares our estimator $\hat{\Phi}_{1,h}$ of the first absolute moment of the mixing density (3.30) with the estimator $\hat{\Phi}_{CL}$ of Cai and Low (2011). We show that, although the two estimators have similar precision, $\hat{\Phi}_{1,h}$ is adaptive to the choice of parameters while $\hat{\Phi}_{CL}$ is sensitive to the choice of the parameter $M_{n,0}$, the maximum of values of $|\theta_i|$,

$i = 1, \dots, n$. In addition, our simulations confirm that the precisions of estimation of functionals Φ and Φ_n differ only insignificantly. Indeed, all results derived in the paper for estimation of a linear functional Φ of the deconvolution density can be automatically applied to estimation of the functional in indirect observations Φ_n .

Acknowledgments

Marianna Pensky was partially supported by National Science Foundation (NSF), grants DMS-1106564 and DMS-1407475. The author wants to thank Alexandre Tsybakov, Oleg Lepski and Alexander Goldenshluger for extremely valuable suggestions and discussions.

References

- [1] Butucea, C., Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, **15**, 69–98.
- [2] Cai, T. T., Jin, J., Low, M. G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.*, **35**, 2421–2449.
- [3] Cai, T., Low, M. (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.*, **31**, 1140–1153.
- [4] Cai, T., Low, M. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.*, **32**, 552–576.
- [5] Cai, T., Low, M. (2005). On adaptive estimation of linear functionals. *Ann. Statist.*, **33**, 2311–2343.
- [6] Cai, T., Low, M. (2011). Testing composite hypotheses, Hermite polynomials, and optimal estimation of a nonsmooth functional. *Ann. Statist.*, **39**, 1012–1041.
- [7] Dattner, I., Goldenshluger, A., and Juditsky, A. (2011) On deconvolution of distribution functions. *Ann. Statist.*, **39**, 2477–2501.
- [8] Delaigle, A. (2007) Nonparametric density estimation from data with a mixture of Berkson and classical errors. *Canadian Journ. Statist.*, **35**, 89–104.
- [9] Donoho, D., Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, **32**, 962–994.
- [10] Ibragimov, I. A., Hasminskii, R. Z. (1984). Nonparametric estimation of the values of a linear functional in Gaussian white noise. *Theory Probab. Appl.*, **29**, 19–32.
- [11] Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, **38**, 481–482.
- [12] Goldenshluger, A., Pereverzev, S.V. (2000) Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Relat. Fields*, **118**, 169–186.
- [13] Gradshteyn, I.S., Ryzhik, I.M. (1980) *Tables of integrals, series, and products*. Academic Press, New York.

- [14] Hall, P., Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.*, **38**, 1686–1732.
- [15] Jaffe, A.H. (2010). Bayesian analysis of cosmic microwave background data. In *Bayesian Methods In Cosmology*, Eds. Hobson, M.P., Jaffe, A.H., Liddle, A.R., Mukherjee, P., Parkinson, D., 2010, Cambridge University Press, New York.
- [16] Klemelä, J., Tsybakov, A. B. (2001). Sharp adaptive estimation of linear functionals. *Ann. Statist.*, **29**, 1567–1600.
- [17] Laurent, B., Ludeña, C., Prieur, C. (2008). Adaptive estimation of linear functionals by model selection. *Electron. J. Stat.*, **2**, 9931020.
- [18] Lepski, O.V. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimators. *Theory of Probability and its Applications*, **36**, 682–697.
- [19] Lepski, O. V., Levit, B. Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.*, **7**, 123–156.
- [20] Lepski, O.V., Mammen, E., Spokoiny, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors *Ann. Statist.*, **25**, 929–947.
- [21] Liang, C.-L., Rice, J.A., de Pater, I., Alcock, C., Axelrod, T., Wang, A., Marshall, S. (2004). Statistical methods for detecting stellar occultations by Kuiper Belt objects: the Taiwanese-American Occultation Survey. *Statist. Sci.*, **19**, 265–274.
- [22] Macdiarmid, J., Blundell, J. (1998). Assessing dietary intake: Who, what and why of under-reporting. *Nutrition Research Reviews*, **11**, 231–253
- [23] Mathé, P., Pereverzev, S.V. (2002). Direct estimation of linear functionals from indirect noisy observations. *Journ. of Complexity*, **18**, 500–516.
- [24] Tsybakov, A. (2009) *Introduction to Nonparametric Estimation*, Springer Series in Statistics.
- [25] Zayed, A.I. (1996) *Handbook of Function and Generalized Function Transformations*. CRC Press, Boca Raton.

7 Proofs

7.1 Proof of the adaptive upper bounds for the risk: the standard case

Proof of Theorem 2. Note that the values \hat{h}_n that deliver optimal (or nearly optimal up to a logarithmic factor of n) are known for all cases except $b = \beta = d = \gamma = 0$, $a < \alpha + 1/2$, can be derived by the straightforward calculus. If $b = \beta = d = \gamma = 0$, $a < \alpha + 1/2$, denote

$$\psi_h(y) = \frac{1}{2\pi} \int_{-h^{-1}}^{h^{-1}} e^{-i\omega y} \frac{\varphi^*(\omega)}{g^*(\omega)} d\omega,$$

$$C_{\psi,1} = \frac{C_{\varphi 2}}{C_{g1}\pi(\alpha - a + 1)}, \quad C_{\psi,2} = \frac{C_{\varphi 2}^2 \|g\|_\infty}{2\pi(2\alpha - 2a + 1) C_{g1}^2}, \quad C_{B,0} = \left(\frac{2s + 2a - 1}{2B^2} \right)^{\frac{1}{2s+2\alpha}}.$$

Observe that $\hat{\Phi}_h$ can be presented as

$$\hat{\Phi}_h = \frac{1}{n} \sum_{l=1}^n \psi_h(Y_l).$$

and also

$$\|\psi_h\|_\infty \leq \frac{C_{\psi,1}}{h^{(\alpha-a+1)}}, \quad \text{Var}(\hat{\Phi}_h) \leq \frac{C_{\psi,2}}{n h^{(2\alpha-2a+1)}}, \quad |\mathbb{E}\hat{\Phi}_h - \Phi| \leq \frac{BC_{\varphi 2} h^{s+a-1/2}}{\sqrt{2\pi(2s+2a-1)}}. \quad (7.1)$$

Then, applying Bernstein inequality to $\sum n^{-1}[\psi_h(Y_l) - \mathbb{E}\hat{\Phi}_h]$ obtain

$$\mathbb{P}\left(|\hat{\Phi}_h - \mathbb{E}\hat{\Phi}_h| > C_\tau n^{-1/2} \sqrt{\log n} h^{a-\alpha-1/2}\right) \leq 2n^{-\tau} \quad (7.2)$$

for any $\tau > 0$ and C_τ satisfying the following inequality

$$C_\tau \geq \frac{2\tau C_{\psi,1}}{3} + \sqrt{\left(\frac{2\tau C_{\psi,1}}{3}\right)^2 + 2\tau C_{\psi,2}}.$$

Set $\tau = 2(2\alpha - 2a + 1)$ and let h_0 and j_0 be such that

$$h_0 = C_{B,0} n^{-\frac{1}{2s+2\alpha}}, \quad h_{j_0} \leq h_0 < h_{j_0-1}.$$

Now, note that by definition (2.14) of \hat{j} , one has

$$\mathbb{E}(\hat{\Phi}_{h_{\hat{j}}} - \Phi)^2 = \mathbb{E}[(\hat{\Phi}_{h_{\hat{j}}} - \Phi)^2 \mathbb{I}(\hat{j} \leq j_0)] + \mathbb{E}[(\hat{\Phi}_{h_{\hat{j}}} - \Phi)^2 \mathbb{I}(\hat{j} > j_0)] \equiv \Delta_1 + \Delta_2. \quad (7.3)$$

Here,

$$\begin{aligned} \Delta_1 &\leq 2\mathbb{E}[(\hat{\Phi}_{h_{\hat{j}}} - \hat{\Phi}_{h_{j_0}})^2 \mathbb{I}(\hat{j} \leq j_0)] + 2\mathbb{E}[(\hat{\Phi}_{h_{j_0}} - \Phi)^2 \mathbb{I}(\hat{j} \leq j_0)] \\ &\leq 2C_{\Phi}^2 n^{-1} \log n h_{j_0}^{-(2\alpha-2a+1)} + 2C_{B,0} n^{-\frac{2s+2a-1}{2s+2\alpha}} \asymp n^{-\frac{2s+2a-1}{2s+2\alpha}} \end{aligned} \quad (7.4)$$

In order to find an upper bound for Δ_2 , denote $D_n(h) = 0.5 C_{\Phi} \sqrt{\log n/n} h^{a-\alpha-1/2} - C_{B,\varphi} h^{s+a-1/2}$, $C_{B,\varphi} = B C_{\varphi 2}/(\pi \sqrt{4s+4a-2})$ and

$$\mathcal{M}_j = \left\{ \omega : \hat{j} = j \right\} = \left\{ \omega : \exists \tilde{j} > j \text{ such that } |\hat{\Phi}_{h_j} - \hat{\Phi}_{h_{\tilde{j}}}| > C_{\Phi} n^{-1/2} \sqrt{\log n} h_j^{-(\alpha-a+1/2)} \right\}.$$

By formula (7.1), one has

$$|\widehat{\Phi}_{h_j} - \widehat{\Phi}_{h_{\tilde{j}}}| \leq |\widehat{\Phi}_{h_j} - \mathbb{E}\widehat{\Phi}_{h_j}| + |\widehat{\Phi}_{h_{\tilde{j}}} - \mathbb{E}\widehat{\Phi}_{h_{\tilde{j}}}| + C_{B,\varphi} \left(h_j^{s+a-1/2} + h_{\tilde{j}}^{s+a-1/2} \right).$$

Since $\tilde{j} > j$, so that $h_{\tilde{j}} < h_j$, derive

$$\mathbb{P}(\mathcal{M}_j) \leq \mathbb{P}(|\widehat{\Phi}_{h_j} - \mathbb{E}\widehat{\Phi}_{h_j}| > D_n(h_j)) + \mathbb{P}(|\widehat{\Phi}_{h_{\tilde{j}}} - \mathbb{E}\widehat{\Phi}_{h_{\tilde{j}}}| > D_n(h_{\tilde{j}}))$$

Moreover, since for $j \geq j_0 + 1$, one has $h_j \leq h_0/2$, assumption (2.16) on C_Φ guarantees that

$$\min \left(D_n(h_j), D_n(h_{\tilde{j}}) \right) \geq C_\tau n^{-1/2} \sqrt{\log n} h^{a-\alpha-1/2},$$

so that, by (7.2), one obtains $\mathbb{P}(\mathcal{M}_j) \leq 4n^{-\tau}$. Finally, since one can show by direct calculations that for $j \leq J$ and some positive absolute constant C , one has

$$\mathbb{E}(\widehat{\Phi}_{h_j} - \Phi)^4 \leq C n^{4\alpha-4a} (\log n)^{4a-4\alpha-2},$$

derive

$$\Delta_2 \leq \sum_{j=j_0+1}^J \sqrt{\mathbb{E}(\widehat{\Phi}_{h_j} - \Phi)^4} \sqrt{\mathbb{P}(\mathcal{M}_j)} \leq C n^{2\alpha-2a-\tau/2} \leq C n^{-1}, \quad (7.5)$$

since $\tau = 2(2\alpha - 2a + 1)$. Combination of formulas (7.3), (7.4) and (7.5) complete the proof.

7.2 Proof of the lower bounds for the risk: the standard case

Proof of Theorem 3. The proof is based on application of Theorems 2.1 and 2.2 of Tsybakov (2009) with the chi-squared divergence. Introduce function

$$K_v^*(\omega) = \begin{cases} (|\omega| - 1)^v \mathcal{P}_1(|\omega|), & 1 < \omega < 2, \\ 1, & 2 \leq |\omega| \leq 3, \\ (4 - |\omega|)^v \mathcal{P}_2(|\omega|), & 3 < \omega < 4, \\ 0, & \text{otherwise} \end{cases} \quad (7.6)$$

where \mathcal{P}_1 and \mathcal{P}_2 are such that $\mathcal{P}_1(z) \neq 0$ for $1 \leq z \leq 2$, $K_v^*(\omega)$ is $(v-1)$ times continuously differentiable on the whole real line (i.e., $\mathcal{P}_1(2) = \mathcal{P}_2(3) = 1$), and $0 \leq K_v^*(\omega) \leq 1$. It is easy to see that function $K_v^*(\omega)$ is even, real-valued and non-negative.

Consider two functions

$$f_1(\theta) = b\pi^{-1}(1 + b^2\theta^2)^{-1}, \quad f_2(\theta) = f_1(\theta) + \lambda f_\delta(\theta) \quad (7.7)$$

where Fourier transform $f_\delta^*(\omega)$ of $f_\delta(\theta)$ is given by

$$f_\delta^*(\omega) = K^*(\omega h) \exp \{i \arg(\varphi^*(\omega))\} \quad (7.8)$$

and function $K^*(\omega) = K_3^*(\omega)$ is of the form (7.6) with $v = 3$. Observe that $f_1(\theta)$ is a pdf. In order to ensure that $f_2(\theta)$ is a pdf, note that, by Assumption on $\rho(\omega) = \arg(\varphi^*(\omega))$ and by formula (7.6), function $f_\delta^*(\omega)$ is twice continuously differentiable and absolutely integrable if $h < \omega_0$, hence, $f_\delta(\theta) \leq C_\delta \theta^{-2}$ as $|\theta| \rightarrow \infty$. Therefore, if λ is small enough, $f_2(\theta) \geq 0$. Moreover, $f_\delta^*(0) = \int_{-\infty}^{\infty} f_\delta(\theta) d\theta = 0$, so that $f_2(\theta)$ is also a pdf.

Choose b in $f_1(\theta)$ small enough, so that $f_1 \in \Omega_s(B/2) \subset \Omega_s(B)$ and let

$$\lambda^2 = C_{bs}^2 h^{2s+1} \quad \text{with} \quad C_{bs}^2 = B^2(s+1)17^{-s}/4. \quad (7.9)$$

Then, direct calculations yield that $\int_{-\infty}^{\infty} \lambda^2 |f_\delta(\omega)|^2 (\omega^2 + 1)^s d\omega \leq B^2/4$, and, therefore, $f_2 \in \Omega_s(B)$.

Now, let us evaluate the difference

$$D_h = |\Phi(f_1) - \Phi(f_2)| = \lambda \left| \int_{-\infty}^{\infty} \varphi^*(-\omega) f_\delta^*(\omega) d\omega \right| = \lambda \int_{-\infty}^{\infty} |\varphi^*(\omega)| |K^*(\omega h)| d\omega. \quad (7.10)$$

Using formula (7.9) and Lemma 3 in Section 7.5 with $A = a/2$, $G = d$, $\aleph = b$, $v = 3$ and $l = 1$, we derive

$$D_h \geq \begin{cases} Ch^{s+a-1/2}, & \text{if } d = 0, \\ Ch^{s+a+4b-1/2} \exp(-dh^{-b}), & \text{if } d > 0 \end{cases} \quad (7.11)$$

Denote $q_k(y) = \int_{-\infty}^{\infty} g(y - \theta) f_k(\theta) d\theta$, and let $Q_k(\mathbf{Y}) = \prod_{i=1}^n q_k(Y_i)$, be the pdf of the sample $Y_1 \dots, Y_n$ under f_k , $k = 1, 2$. We shall find a combination of parameters n and h which ensures that the chi-squared divergence $\chi^2(Q_1, Q_2)$ between Q_1 and Q_2 , is bounded above. Then, by Theorems 2.1 and 2.2 of Tsybakov (2009), one has $R_n(\Omega_s(B)) \geq D_h^2$ where D_h is given by (7.11).

By Lemma 2, one has $q_1(x) \geq C_{bg}(x^2 + 1)^{-1}$, and Lemma 4 implies that, in order to ensure $\chi^2(Q_1, Q_2) \leq \kappa^2$, it is sufficient to establish that

$$H = \int_{-\infty}^{\infty} (x^2 + 1) [q_2(x) - q_1(x)]^2 dx \leq C_{bg} n^{-1} \log(1 + \kappa^2) \quad (7.12)$$

for some $\kappa \in (0, 1)$. Note that H in (7.12) can be written as

$$H = \lambda^2 \int_{-\infty}^{\infty} (x^2 + 1) [q_\delta(x)]^2 dx.$$

Then, H can be split as $H = \lambda^2(H_1 + H_2)$ where

$$H_1 = \int_{-\infty}^{\infty} [q_\delta(x)]^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |q_\delta^*(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |g^*(\omega)|^2 |K^*(\omega h)|^2 d\omega$$

and

$$H_2 = \int_{-\infty}^{\infty} x^2 [q_\delta(x)]^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{dq_\delta^*(\omega)}{d\omega} \right|^2 d\omega \leq H_{21} + h^2 H_{22} + H_{23}$$

with

$$\begin{aligned} H_{21} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |(g^*)'(\omega)|^2 |K^*(\omega h)|^2 d\omega \\ H_{22} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |g^*(\omega)|^2 |(K^*)'(\omega h)|^2 d\omega \end{aligned}$$

and $H_{23} = \rho^2 H_1$. Taking into account condition (2.19) and applying Lemma 3 with $A = \alpha$, $G = 2\gamma$, $\aleph = \beta$, $l = 2$ and $v = 3$ for H_1 , with $A = \alpha - \tau$, $G = 2\gamma$, $\aleph = \beta$, $l = 2$ and $v = 3$ for H_{21} and $A = \alpha$, $G = 2\gamma$, $\aleph = \beta$, $l = 2$ and $v = 2$ for H_{22} , obtain that

$$H \leq \begin{cases} C\lambda^2 h^{2\alpha-1}, & \text{if } \gamma = 0, \\ C\lambda^2 h^{2\alpha+U_\tau} \exp(-2\gamma h^{-\beta}), & \text{if } \gamma > 0. \end{cases} \quad (7.13)$$

Due to the choice of λ given by (7.9), combination of formulae (7.12) and (7.13) yield the following expression for $h = h(n)$:

$$h = \begin{cases} Cn^{-\frac{1}{2\alpha+2s}}, & \text{if } \gamma = 0, \\ [(2\gamma)^{-1} (\log n - (2s + 2\alpha + U_\tau + 1) \log \log n)]^{-1/\beta}, & \text{if } \gamma > 0. \end{cases}$$

In order to complete the proof, recall that, by Theorems 2.1 and 2.2 of Tsybakov (2009), one has $R_n(\Omega_s(B)) \geq D_h^2$ where D_h is given by (7.11).

Finally, consider the case when convergence rates are paarmetric. In this situation, just set $\lambda = n^{-1/2}$ and choose $f_1, f_\delta \in \Omega_s(B/2)$ such that f_δ integrates to zero and $f_2 = f_1 + \lambda f_\delta$ is nonnegative. Then, the lower bounds can be obtained directly from Theorems 2.1 and 2.2 of Tsybakov (2009).

7.3 Proofs of statements in Section 3

Proof of Theorem 4. For derivation of (3.27), observe that $\mathbb{E}\widehat{\Phi}_1 = \Phi_1$ and, moreover, due to assumption A1 and conditions (3.24) – (3.26), one has $\text{Var}(\widehat{\Phi}_1) \leq Cn^{-1}$. For the same reason, the values of $A_{m,j,k,l}$, $l = 1, 2$, are uniformly bounded, so that $\mathbb{E}\widehat{F_m}(1) = F_m(1)$ and $\text{Var}[\widehat{F_m}(1)] \leq Cn^{-1}$. Observe that under assumption (3.25), one has

$$\begin{aligned} \text{Var}[\widehat{\Phi}_{20h}] &\asymp n^{-1} \int_1^{h^{-1}} (\omega^2 + 1)^{\alpha-(m+1)a_m} \exp(-2d\omega^b + 2\gamma\omega^\beta) d\omega \\ \mathbb{E}\widehat{\Phi}_{20h} - \Phi_{20} &\asymp \int_{h^{-1}}^\infty (\omega^2 + 1)^{-(m+1)a_m} \exp(-2d\omega^b) |f^*(\omega)| d\omega \end{aligned}$$

In order to complete the proof, note that $\mathbb{E}\widehat{\Phi}_h - \Phi = \mathbb{E}\widehat{\Phi}_{20h} - \Phi_{20}$ and $\text{Var}[\widehat{\Phi}_h] \leq \text{Var}[\widehat{\Phi}_{20h}] + Cn^{-1}$ and use the Cauchy inequality if $f \in \Omega_s(B)$ or upper bounds for $|f^*|$ if $f \in \Xi_s(B)$.

Proof of Theorem 5. Observe that under assumptions of Theorem 5, conditions of Theorem 4 are satisfied with $m = 2M + 1$, $a_m = 1$ and $b = d = 0$. Hence, inequality (3.27) is valid. To complete the proof of the upper bounds (3.39) for the minimax risk, use formula (2.11) in Lemma 1 with $A_1 = 2M + 1 + s$, $A_2 = \alpha - (2M + 2)$ and $b = d = 0$.

In order to establish the lower bounds (3.40) for the minimax risk, use the methodology suggested by Theorems 2.1 and 2.2 of Tsybakov (2009). Similarly to the proof of Theorem 3, choose

$$f_1(\theta) = \tilde{C} b(1 + b^2\theta^2)^{-(2M+3)}, \quad f_2(\theta) = f_1(\theta) + \lambda f_\delta(\theta), \quad (7.14)$$

where $f_\delta(\theta) = h^{-1}K_v(\theta/h)$, $f_\delta^*(\omega) = K_v^*(\omega h)$ with $v = 4M + 4M + 7$ and $K_v^*(\omega)$ is defined by formula (7.6). Note that function $f_\delta^*(\omega)$ is $(4M + 6)$ times continuously differentiable and absolutely integrable, hence, $f_\delta(\theta) \leq C_\delta \theta^{-(4M+6)}$ as $|\theta| \rightarrow \infty$. Moreover, $f_\delta^*(0) = \int_{-\infty}^\infty f_\delta(\theta) d\theta = 0$. Therefore, if λ is small enough, $f_2(\theta) \geq 0$, so that $f_2(\theta)$ is also a pdf.

Let B_1, B_2 and b be such that $f_1 \in \Omega_s(B_1/2, B_2/2)$. Note that

$$\begin{aligned} \int_{-\infty}^\infty \theta^{4M+4} f(\theta) d\theta &\leq \int_{-\infty}^\infty \theta^{4M+4} f_1(\theta) d\theta + \lambda \int_{-\infty}^\infty \theta^{4M+4} h^{-1} K(\theta/h) d\theta \\ &\leq B_1/2 + \lambda h^{4M+4} \int_{-\infty}^\infty z^4 K(z) dz \leq B_1 \end{aligned}$$

if λ and h are small enough. Also, by definition of $K_v^*(\omega)$, one has $|K_v^*(\omega)| \leq 1$ and $|\omega|^s |K_v^*(\omega)| \leq 4^s$, thus

$$\lambda |f_\delta^*(\omega)| (|\omega|^s + 1) = \lambda |K^*(\omega h)| [|\omega h|^s h^{-s} + 1] \leq \lambda (4^s h^{-s} + 1) \leq B_2/2$$

provided $\lambda = C_{sB} h^s$ for some absolute constant C_{sB} that depends on s and B_2 only. Now, using integration by part and recalling that $f_\delta^*(\omega) = K_v^*(\omega h)$, one derives

$$\begin{aligned} D_h &= |\Phi(f_1) - \Phi(f_2)| = 2\lambda\pi^{-1}(2M+1)! \left| \int_{-\infty}^{\infty} \omega^{-(2M+2)} f_\delta^*(\omega) d\omega \right| \\ &= 2\lambda\pi^{-1}(2M+1)! h^{2M+1} \int_1^4 z^{-(2M+2)} K_v^*(z) dz \geq C\lambda h^{2M+1} \geq C h^{s+2M+1}. \end{aligned} \quad (7.15)$$

Finally, similarly to the proof of Theorem 3, due to Lemma 2, one has $q_1(x) \geq C_{bg}(x^2 + 1)^{-(2M+3)}$, and Lemma 4 implies that it is sufficient to ensure that

$$H = \int_{-\infty}^{\infty} (x^2 + 1)^{2M+3} [q_1(x) - q_2(x)]^2 dx = \lambda^2 \int_{-\infty}^{\infty} (x^2 + 1)^{2M+3} [q_\delta(x)]^2 dx \leq C_{bg} n^{-1} \log(1 + \kappa^2). \quad (7.16)$$

Following the steps in the proof of Theorem 3, we construct an upper bound of H in (7.16) as $H \leq 2^{2M+2} \lambda^2 (H_1 + H_2)$ with

$$H_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |g^*(\omega)|^2 |K_v^*(\omega h)|^2 d\omega \asymp h^{2\alpha-1+\beta(2v+1)} \exp(-2\gamma h^{-\beta})$$

and

$$H_2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{d^{2M+3} q_\delta^*(\omega)}{d\omega^{2M+3}} \right|^2 d\omega \leq \sum_{k=0}^{2M+3} \int_{-\infty}^{\infty} \left| \frac{d^{2M+3-k} g^*(\omega)}{d\omega^{2M+3-k}} \right|^2 \left| \frac{d^k K_v^*(\omega h)}{d\omega^k} \right|^2 d\omega \quad (7.17)$$

$$\asymp \sum_{k=0}^{2M+3} h^{2k} \int_{-\infty}^{\infty} (\omega^2 + 1)^{-\alpha+\tau(2M+3-k)} \exp(-2\gamma|\omega|^\beta) |(K_v^*)^{(k)}(\omega h)| d\omega. \quad (7.18)$$

If $\gamma = \beta = 0$, then $H_2 \asymp h^{2\alpha-1}$. Otherwise, apply Lemma 3 with $v = 4M + 7 - k$ for the k -th term in (7.18). Obtain $H_2 \asymp h^{2\alpha+\beta-1+\vartheta} \exp(-2\gamma h^{-\beta})$ where $\vartheta = \min(4M + 6 + \beta(4M + 8), \beta(4M + 15) - \tau(4M + 6))$. Hence, derive

$$H \leq \begin{cases} C\lambda^2 h^{2\alpha-1}, & \text{if } \gamma = 0, \\ C\lambda^2 h^{2\alpha+\beta-1+U_\tau^*} \exp(-2\gamma h^{-\beta}), & \text{if } \gamma > 0. \end{cases} \quad (7.19)$$

where $U_\tau^* = \min(4M + 6 + \beta(4M + 8), \beta(8M + 15) - \tau(4M + 6), \beta(8M + 14))$. With $\lambda = C_{sB} h^s$, (7.19) implies that inequality (7.16) holds provided $h = Cn^{-1/(2s+2\alpha-1)}$ if $\gamma = 0$, $\beta = 0$, and $h = [(2\gamma)^{-1}(\log n - (2\alpha + \beta - 1 + U_\tau^*) \log \log n)]^{-1/\beta}$ if $\gamma > 0$, $\beta > 0$. In order to complete the proof, recall that, by Theorems 2.1 and 2.2 of Tsybakov (2009), one has $R_n(\Omega_s(B)) \geq D_h^2$ where D_h is given by (7.15).

7.4 Proofs of the statements in Section 4

Proof of Theorem 7. Let $\mu_n \leq 1/2$. Otherwise, $k_n \geq n/2$ and there is no point considering the case as being sparse. Consider two mixing pdfs

$$f_{\mu,k}(\theta) = (1 - \mu_n)\delta(\theta) + \lambda\mu_n f_k(\theta), \quad k = 1, 2, \quad (7.20)$$

and corresponding marginal densities

$$q_{\mu,k}(x) = (1 - \mu_n)g(x) + \lambda\mu_n q_k(x), \quad k = 1, 2, \quad (7.21)$$

where $q_k(y) = \int_{-\infty}^{\infty} g(y - \theta) f_k(\theta) d\theta$ and $q_{\mu,k}(y) = \int_{-\infty}^{\infty} g(y - \theta) f_{\mu,k}(\theta) d\theta$. Again, as in the proof of Theorem 3, let $Q_{\mu,k}(\mathbf{Y}) = \prod_{i=1}^n q_{\mu,k}(Y_i)$, $k = 1, 2$, be the pdf of the sample $Y_1 \cdots, Y_n$ under f_k , $k = 1, 2$. Then, by Lemma 4, one has $\chi^2(Q_{\mu,1}, Q_{\mu,2}) \leq \kappa^2$ provided

$$H_{\mu} = \int_{-\infty}^{\infty} [q_{\mu 1}(x)]^{-1} [q_{\mu 1} - q_{\mu 2}]^2 dx \leq n^{-1} \log(1 + \kappa^2). \quad (7.22)$$

Direct calculations yield that

$$\begin{aligned} H_{\mu} &\leq \lambda^2 \mu_n^2 \int_{-\infty}^{\infty} [(1 - \mu_n)g(x) + \lambda\mu_n q_1(x)]^{-1} \left[\int_{-\infty}^{\infty} g(x - \theta)(f_2(\theta) - f_1(\theta)) d\theta \right]^2 dx \\ &\leq \frac{\lambda^2 \mu_n^2}{(1 - \mu_n)^2} (I_1 + I_2) \leq 8C_I \lambda^2 \mu_n^2 \end{aligned}$$

since $\mu_n \geq 1/2$ and $(x - y)^2 \leq x^2 + y^2$ for $x, y \geq 0$. Therefore, $H_{\mu} \leq n^{-1} \log(1 + \kappa^2)$ provided $\lambda^2 = \lambda_0^2 \min(1, (n\mu_n^2)^{-1})$ where $\lambda_0^2 = \log(1 + \kappa^2)/(8C_I)$. Finally, due to (4.9), one has

$$D_h^2 = [\Phi(f_{\mu 1}) - \Phi(f_{\mu 2})]^2 = \lambda^2 \Delta_{12}^2 = \lambda_0^2 \Delta_{12}^2 \min(1, (n\mu_n^2)^{-1})$$

which, together with Theorems 2.1 and 2.2 of Tsybakov (2009), completes the proof.

Proof of Corollary 4. Before all else, observe that if $\varphi(x)$ is constant, $\varphi(x) = C$ for every x , then $\Phi_{\mu} = C$ and estimation is unnecessary.

First, consider the case when the even part of $\varphi(x)$, $\varphi_{\text{even}}(x) = 0.5(\varphi(x) + \varphi(-x))$ is not identically equal to zero. Choose two values $\rho_1, \rho_2 \in (0, 1)$, $\rho_1 \neq \rho_2$ and set $f_k(\theta) = f(\theta|\rho_k)$ where $f(\theta|\rho) = \mathcal{N}(\theta|0, \sigma^2\rho)$, $k = 1, 2$, the Gaussian pdf with zero mean and variance $\sigma^2\rho_k$. Then, direct calculations yield that $q_k(x) = \mathcal{N}(x|0, \sigma^2(1 + \rho_k))$, $k = 1, 2$. Hence,

$$I_k = \int_{-\infty}^{\infty} g^{-1}(x) q_k^2(x) dx = \frac{1}{\sqrt{2\pi(1 + \rho_k)}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2(1 - \rho_k)}{2\sigma^2(1 + \rho_k)}\right) dx, \quad k = 1, 2,$$

and inequality (4.8) holds with $C_I = \max\{(1 - \rho_1)^{-1/2}, (1 - \rho_2)^{-1/2}\}$. Furthermore, note that $A_k = G(\rho_k)$ where

$$G(\rho) = \int_{-\infty}^{\infty} \varphi_{\text{even}}(\theta) f(\theta|\rho) d\theta = \frac{2}{\sigma\sqrt{2\pi\rho}} \int_0^{\infty} z^{-1/2} \varphi(\sqrt{z}) e^{-\frac{z}{2\sigma^2\rho}} dz.$$

Inequality (4.9) is violated only if $G(\rho)$ takes constant value for $0 < \rho < 1$. It is easy to notice that $\sqrt{\rho} G(\rho)$ is proportional to the Laplace transform of the function $z^{-1/2} \varphi_1(\sqrt{z})$. Hence, $G(\rho)$ is constant if and only if the Laplace transform of $z^{-1/2} \varphi_1(\sqrt{z})$ is equal to $Cz^{-1/2}$, which is possible only if $\varphi(x)$ is a constant function. Therefore, if $\varphi_{\text{even}}(x)$ does not vanish, it is always possible to choose two values, $\rho_1 \neq \rho_2$ in $(0, 1)$ such that inequality (4.9) holds.

Now, consider the situation when $\varphi_{\text{even}}(x) \equiv 0$. Then, $\varphi(\theta)$ is an odd function. Choose $\rho \in (0, 1)$ and set $f_1(\theta) = f(\theta|\rho) = \mathcal{N}(\theta|0, \sigma^2\rho)$. Let $f_2(\theta) = f_1(\theta) + f_{\delta}(\theta)$ where

$$f_{\delta}(\theta) = \text{sign}(\theta) f_1(\theta).$$

Since, $|f_\delta(\theta)| \leq f_1(\theta)$ for any θ , $f_2(\theta) \geq 0$. Moreover, condition (4.8) holds for $k = 1, 2$. Furthermore, since $f_\delta(\theta)$ integrates to zero, $f_2(\theta)$ is a pdf. It remains to check that

$$\int_{-\infty}^{\infty} \varphi(\theta) f_\delta(\theta) d\theta = 2 \int_0^{\infty} \varphi(\theta) f(\theta|\rho) d\theta \neq 0$$

for some $\rho \in (0, 1)$. The latter can be accomplished in the same manner as in the case when $\varphi_{\text{even}}(\theta) \neq 0$.

Proof of Theorem 8. First, let us prove parametric rates. For this purpose, we choose f_1 and f_2 in Theorem 7 such that $f_k^*(\omega)$ are ς_0 times continuously differentiable with

$$\left| \frac{d^l f_k^*(\omega)}{d\omega^l} \right| \leq C_f |f_k^*(\omega)|, \quad l = 1, 2, \dots, \varsigma_0, \quad \int_{-\infty}^{\infty} (1 + |\omega|^{2\tau\varsigma_0}) |f_k^*(\omega)|^2 d\omega < \infty, \quad k = 1, 2. \quad (7.23)$$

Let $f_1^*(\omega)$ and $f_2^*(\omega)$ also satisfy the condition (4.9), i.e.

$$\Delta_{12} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [f_1^*(\omega) - f_2^*(\omega)] \varphi^*(-\omega) d\omega \neq 0.$$

It is always possible to find functions f_1 and f_2 like this. In order to prove parametric rates (4.10), it is sufficient to show that inequalities (4.8) hold. For this purpose, note that $g^{-1}(x) \leq C_{g1}^{-1}(x^2 + 1)^\varsigma \leq C_\varsigma C_{g1}^{-1}(x^{2\varsigma} + 1)$, where constant C_ς depends on ς only. Hence,

$$I_k = \int_{-\infty}^{\infty} g^{-1}(x) \left[\int_{-\infty}^{\infty} g(x - \theta) f_k(\theta) d\theta \right]^2 dx \leq C_\varsigma C_{g1}^{-1} (I_{k1} + I_{k2}),$$

where

$$I_{k1} = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x - \theta) f_k(\theta) d\theta \right]^2 dx \leq \int_{-\infty}^{\infty} g(x - \theta) dx \int_{-\infty}^{\infty} g(x - \theta) f_k^2(\theta) d\theta dx < \infty$$

and

$$\begin{aligned} I_{k2} &= \int_{-\infty}^{\infty} |x|^{2\varsigma} \left[\int_{-\infty}^{\infty} g(x - \theta) f_k(\theta) d\theta \right]^2 dx \leq \int_{-\infty}^{\infty} \left[x^{\varsigma_0} \int_{-\infty}^{\infty} g(x - \theta) f_k(\theta) d\theta \right]^2 dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\frac{d^{\varsigma_0} [g^*(\omega) f_k^*(\omega)]}{d\omega^{\varsigma_0}} \right]^2 d\omega < \infty \end{aligned}$$

due to conditions (7.23). Therefore, due to Theorem 7, the first three inequalities in (4.13) are valid.

Now, we need to prove nonparametric rates in (4.13). Let, similarly to the proof of Theorem 7, $f_{\mu,k}(\theta)$ and $q_{\mu,k}(x)$ be defined in (7.20) and (7.21), respectively, where $q_k(y) = \int_{-\infty}^{\infty} g(y - \theta) f_k(\theta) d\theta$ and $q_{\mu,k}(y) = \int_{-\infty}^{\infty} g(y - \theta) f_{\mu,k}(\theta) d\theta$. Similarly to the proof of Theorem 3, we consider $f_1(\theta) = b\pi^{-1}(1 + b^2\theta^2)^{-1}$, so that $f_1(\theta) \in \Omega_s(B/4)$ if b is small enough. Then, by Lemma 2, one has $q_1(x) \geq C_{bg}(x^2 + 1)^{-1}$. Let $f_2(\theta) = f_1(\theta) + \lambda f_\delta(\theta)$ where Fourier transform $f_\delta^*(\omega)$ of $f_\delta(\theta)$ is given by (7.8) with $K^*(\omega) = K_{2\varsigma_0+1}^*(\omega)$ where function $K_v^*(\omega)$ is defined by (7.6). Again, since f_δ^* is $(2\varsigma_0)$ times continuously differentiable for $h < \omega_0^{-1}$ and is absolutely integrable, one has $f_\delta(\theta) \leq C_\delta |\theta|^{-2\varsigma_0}$ as $|\theta| \rightarrow \infty$. Therefore, if λ is small enough, $f_2(\theta) \geq 0$ and it is a pdf. Also, similarly to the proof of Theorem 7, we choose λ^2 of the form (7.9), so that $f_2 \in \Omega_s(B)$.

Now, we need to evaluate the difference $D_h = |\Phi(f_1) - \Phi(f_2)|$ given by (7.10). Using expression for λ in formula (7.9) and Lemma 3 with $A = a/2$, $G = d$, $\aleph = b$, $v = 2\varsigma_0 + 1$ and $l = 1$, we derive

$$D_h \geq \begin{cases} Ch^{s+a-1/2}, & \text{if } d = 0, \\ Ch^{s+a+2b(\varsigma_0+1)-1/2} \exp(-dh^{-b}), & \text{if } d > 0. \end{cases} \quad (7.24)$$

It remains to find relation between n and h which ensures that the chi-squared divergence $\chi^2(Q_{\mu_1}, Q_{\mu_2})$ is bounded above. Lemma 4 implies that $\chi^2(Q_{\mu_1}, Q_{\mu_2}) \leq \kappa^2$ is guaranteed by

$$H_\mu = \lambda^2 \mu_n^2 \int_{-\infty}^{\infty} [g(x)]^{-1} (q_d^*(x))^2 dx \leq n^{-1} \log(1 + \kappa^2). \quad (7.25)$$

Similarly to the proof of Theorem 3, we note that

$$H_\mu \leq C_{\varsigma_0} \lambda^2 \mu_n^2 (H_{\varsigma_0 1} + H_{\varsigma_0 2})$$

where constant C_{ς_0} depends on ς_0 only and, similarly to the proof of Theorem 3,

$$H_{\varsigma_0 1} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |q^*(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |g^*(\omega)|^2 |K^*(\omega h)|^2 d\omega$$

and

$$H_{\varsigma_0 2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{d^{\varsigma_0} q^*(\omega)}{d\omega^{\varsigma_0}} \right|^2 d\omega \leq H_{\varsigma_0 21} + h^{2\varsigma_0} H_{\varsigma_0 22} + H_{\varsigma_0 23}$$

with

$$\begin{aligned} H_{\varsigma_0 21} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{d^{\varsigma_0} g^*(\omega)}{d\omega^{\varsigma_0}} \right|^2 |K^*(\omega h)|^2 d\omega \\ H_{\varsigma_0 22} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |g^*(\omega)|^2 |(K^*)^{(\varsigma_0)}(\omega h)|^2 d\omega \end{aligned}$$

and $H_{\varsigma_0 23} = \rho^2 H_{\varsigma_0 1}$.

Take into account condition (4.12) and apply Lemma 3 with $G = 2\gamma$, $\aleph = \beta$, $l = 2$. Let $A = \alpha$ and $v = 2\varsigma_0 + 1$ for $H_{\varsigma_0 1}$ and $H_{\varsigma_0 23}$, $A = \alpha - \varsigma_0 \tau$, and $v = 2\varsigma_0 + 1$ for $H_{\varsigma_0 21}$ and $A = \alpha$ and $v = \varsigma_0 + 1$ for $H_{\varsigma_0 22}$. Then,

$$H_\mu \leq \begin{cases} C\lambda^2 \mu_n^2 h^{2\alpha-1}, & \text{if } \gamma = 0, \\ C\lambda^2 \mu_n^2 h^{2\alpha+U_{\tau, \varsigma_0}+2s+1} \exp(-2\gamma h^{-\beta}), & \text{if } \gamma > 0. \end{cases} \quad (7.26)$$

Due to (7.9), combination of (7.25), (7.26) and $\mu_n = n^{\nu-1}$ yields the following expression for $h = h(n)$:

$$h = \begin{cases} Cn^{-\frac{2\nu-1}{2\alpha+2s}}, & \text{if } \gamma = 0, \\ \left[\frac{2\nu-1}{2\gamma} \left(\log n - \frac{2\alpha+2s+U_{\tau, \varsigma_0}+1}{2\nu-1} \log \log n \right) \right]^{-1/\beta}, & \text{if } \gamma > 0. \end{cases}$$

In order to complete the proof, recall that, by Theorems 2.1 and 2.2 of Tsybakov (2009), one has $R_n(\Omega_s(B)) \geq D_h^2$ where D_h is given by (7.24).

Proof of Theorem 9. Theorem 9 can be proved by combination of methods used in the proofs of Theorems 3 and 8. Let $f_{\mu, k}(\theta)$ and $q_{\mu, k}(x)$ be defined in (7.20) and (7.21), respectively,

where $q_k(y) = \int_{-\infty}^{\infty} g(y - \theta) f_k(\theta) d\theta$ and $q_{\mu,k}(y) = \int_{-\infty}^{\infty} g(y - \theta) f_{\mu,k}(\theta) d\theta$. Similarly to the proof of Theorem 3, we consider $f_1(\theta) = b\pi^{-1}(1 + b^2\theta^2)^{-1}$, so that $f_1(\theta) \in \Omega_s(B/4)$ if b is small enough. Then, by Lemma 2, one has $q_1(x) \geq C_{bg}(x^2 + 1)^{-1}$. Let $f_2(\theta) = f_1(\theta) + \lambda f_\delta(\theta)$ where Fourier transform $f_\delta^*(\omega)$ of $f_\delta(\theta)$ is given by (7.8) with $K^*(\omega) = K_3^*(\omega)$, where function $K_v^*(\omega)$ is given by (7.6). Then, both f_1 and f_2 are pdfs and $D_h = |\Phi(f_1) - \Phi(f_2)| \geq Ch^{s+a-1/2}$. On the other hand, inequality (7.22) holds whenever, similarly to (7.12), one ensures that

$$H_\mu = \lambda^2 \mu_n C_{bg}^{-1} \int_{-\infty}^{\infty} (x^2 + 1) (q_\delta^*(x))^2 dx \leq n^{-1} \log(1 + \kappa^2).$$

Since $\mu_n = n^{\nu-1}$, the latter is guaranteed by choosing

$$h = [(2\gamma)^{-1} (\nu \log n - (2s + 2\alpha + U_\tau + 1) \log \log n)]^{-1/\beta}$$

and yields $D_h \geq C (\log n)^{-\frac{2s+2\alpha-1}{\beta}}$. Application of Theorems 2.1 and 2.2 of Tsybakov (2009) completes the proof.

7.5 Proofs of the supplementary statements used in Section 7.2

Lemma 2 *Let $\varrho \geq 1$ and $f(\theta) = C_\varrho b(1 + b^2\theta^2)^{-\varrho}$ where C_ϱ depends on ϱ only. Let $A > 0$ be such that*

$$\int_{-A}^A g(x) dx = C_A > 0.$$

Then

$$q(x) = \int_{-\infty}^{\infty} g(x - \theta) f(\theta) d\theta \geq C_{bg}(x^2 + 1)^{-\varrho} \quad (7.27)$$

for some positive constant C_{bg} which depends on ϱ , C_ϱ , b , A and C_A .

Proof of Lemma 2. Observe that

$$q(x) \geq C_\varrho b \int_{-A}^A (1 + b^2(x - z)^2)^{-\varrho} g(z) dz$$

and consider cases $|x| \leq A$ and $|x| > A$ separately. If $|x| \leq A$, then, due to $|x - z| \leq 2A$ and $(x^2 + 1)^{-1} < 1$, one obtains

$$q(x) \geq C_\varrho b (1 + 4b^2 A^2)^{-\varrho} \int_{-A}^A g(z) dz \geq C_\varrho C_A b (1 + 4b^2 A^2)^{-\varrho} \geq C_\varrho C_A b (1 + 4b^2 A^2)^{-\varrho} (x^2 + 1)^{-\varrho}. \quad (7.28)$$

If $|x| > A$, then, due to $|x - z| \leq 2x$, one derives

$$q(x) \geq C_\varrho b (1 + 4b^2 x^2)^{-\varrho} \int_{-A}^A g(z) dz \geq C_\varrho C_A b (1 + 4b^2)^{-\varrho} (1 + x^2)^{-\varrho}. \quad (7.29)$$

In order to obtain (7.27), combine (7.28) and (7.29) and use $C_{bg} = C_\varrho C_A b \max((1 + 4b^2 A^2)^{-\varrho}, (1 + 4b^2)^{-\varrho})$.

Lemma 3 Let function $K_v^*(\omega)$ be given by (7.6) where \mathcal{P}_1 and \mathcal{P}_2 are such that $\mathcal{P}_1(z) \neq 0$ for $1 \leq z \leq 2$, $K_v^*(\omega)$ is $(v-1)$ times continuously differentiable on the whole real line (i.e., $\mathcal{P}_1(2) = \mathcal{P}_2(3) = 1$), and $0 \leq K_v^*(\omega) \leq 1$. Consider

$$J_l = \int_{-\infty}^{\infty} (\omega^2 + 1)^{-A} \exp\left(-G|\omega|^{\aleph}\right) |K_v^*(\omega h)|^l d\omega, \quad l = 1, 2. \quad (7.30)$$

where A , G and \aleph are nonnegative constants such that $A > 0$ and $\aleph = 0$ whenever $G = 0$. Then, as $h \rightarrow 0$, one has

$$J_l \asymp \begin{cases} h^{2A-1}, & \text{if } G = 0, \\ h^{2A+\aleph(lv+1)-1} \exp(-Gh^{-\aleph}), & \text{if } G > 0. \end{cases} \quad (7.31)$$

Proof of Lemma 3. It is easy to see that function $K_v^*(\omega)$ is even, real-valued and non-negative. Re-write J_l as

$$J_l = \frac{2}{h} \int_1^4 (\omega^2 h^{-2} + 1)^{-A} \exp\left(-Gh^{-\aleph} \omega^{\aleph}\right) |K_v^*(\omega)|^l d\omega$$

and note that

$$J_l \asymp h^{2A-1} U_{hl} \quad \text{with} \quad U_{hl} = \int_1^4 \exp\left(-Gh^{-\aleph} \omega^{\aleph}\right) |K_v^*(\omega)|^l d\omega. \quad (7.32)$$

If $G = 0$, then U_{hl} in (7.32) does not depend on h and $J_l \asymp h^{2A-1}$. If $G > 0$, the main contribution in (7.32) comes from interval $(1, 2)$, so that, introducing new variable $z = \omega - 1$, obtain

$$U_{hl} \approx \int_0^1 \exp\left(-Gh^{-\aleph}(z+1)^{\aleph}\right) z^{lv} \mathcal{P}_1(z+1) dz = \int_0^1 \exp(-u_h(z)) z^{lv} \mathcal{P}_1(z+1) dz$$

where function $u_h(z) = Gh^{-\aleph}(z+1)^{\aleph}$ has its minimum on the interval $(0, 1)$ at $z = 0$. Note that $u_h(0) = Gh^{-\aleph}$ and $u'_h(z) = \aleph Gh^{-\aleph} (z+1)^{\aleph-1}$. Since $\aleph(z+1)^{\aleph-1}$ lies between \aleph and $\aleph 2^{\aleph-1}$, the mean value theorem implies that

$$U_{hl} \asymp \exp(-Gh^{-\aleph}) \int_0^1 z^{lv} \mathcal{P}_1(z+1) \exp(-C_{\aleph} Gh^{-\aleph} z) dz,$$

where the constant C_{\aleph} lies between \aleph and $\aleph 2^{\aleph-1}$. Change variables $t = C_{\aleph} Gh^{-\aleph} z$ and recall that \mathcal{P}_1 is a continuous function with $\mathcal{P}_1(x) \neq 0$ for $x \in [1, 2]$, so that $\mathcal{P}_1(x)$ is uniformly bounded above and below for $x \in [1, 2]$ and $\mathcal{P}_1(z+1) \asymp 1$ for $z \in [0, 1]$. Therefore,

$$U_{hl} \asymp \exp(-Gh^{-\aleph}) h^{lv+1} \int_0^{C_{\aleph} Gh^{-\aleph}} t^{lv} \exp(-t) dt \asymp \exp(-Gh^{-\aleph}) h^{lv+1}. \quad (7.33)$$

Combination of (7.32) and (7.33) complete the proof.

Lemma 4 Let $\mathbf{x} = (x_1, \dots, x_n)$ and let $Q_k(\mathbf{x}) = \prod_{i=1}^n q_k(x_i)$, $k = 1, 2$, be two pdfs. Denote

$$\chi^2(Q_1, Q_2) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_1^{-1}(\mathbf{x}) [Q_2(\mathbf{x}) - Q_1(\mathbf{x})]^2 d\mathbf{x},$$

the chi-squared divergence between Q_1 and Q_2 . Then, for any $\kappa \in (0, 1)$

$$\int_{-\infty}^{\infty} [q_1(x)]^{-1} [q_1(x) - q_2(x)]^2 dx \leq n^{-1} \log(1 + \kappa^2) \quad \text{implies} \quad \chi^2(Q_1, Q_2) \leq \kappa^2. \quad (7.34)$$

Proof of Lemma 4. Observe that

$$I_q = \int_{-\infty}^{\infty} [q_1(x)]^{-1} [q_1(x) - q_2(x)]^2 dx = \int_{-\infty}^{\infty} [q_1(x)]^{-1} [q_2(x)]^2 dx - 1.$$

Therefore,

$$\begin{aligned} \chi^2(Q_1, Q_2) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_1^{-1}(\mathbf{x}) Q_2^2(\mathbf{x}) d\mathbf{x} - 1 = \left[\int_{-\infty}^{\infty} [q_1(x)]^{-1} [q_2(x)]^2 dx \right]^n - 1 \\ &= (I_q + 1)^n - 1 = \exp[n \log(I_q + 1)] - 1. \end{aligned}$$

Taking into account that $\log(1 + z) \leq z$ for $0 < z < 1$, one obtains

$$\chi^2(Q_1, Q_2) \leq \exp(nI_q) - 1 \leq \exp[\log(1 + \kappa^2)] - 1 = \kappa^2,$$

which proves (7.34).